


```
QY 142 SVTHIPAAQIPALIAAAYSHYTSYSDSHVFLIYEGVALVLLFNLFALKRNPVWHEIGFGL 201
Db 125 TGGLSLSTLGGIIGLYPLFEFORHOFIYIATINIAFAINAFNMVGLPLVTKSAFIW 184
QY 202 TIALFVISTAILARSNPKAPNSQ-VMTAMSNYTGSDGVCFFILGLSTSCFMFGLDAAM 260
Db 185 SLTGFFVISTILSCASPDYNSGKFVGEFINTGWDGVAWLLGLLOGGFLTGFDGVA 244
QY 261 HLABECTDAARTVPKAVVSAAIIIGCTAFPTYTIAVLYGTDLSILSSAGYIPETWTQS 320
Db 245 HMTIEEPNPTVGLPRIMIGVCIGITGLIFLIVLLFVAGDINLVIESAA-----TPL 297
QY 321 LRSLSFATVLSGGI-----VMAFALNAVOETASRLTWSFARDNGLVFSHTLERHP 373
Db 298 LQTHDATNSNAGSICLLMPLPCTVFAATIMTSSRMVAFARDCGLPASPFSSRVHP 357
QY 374 RMQVPVMSLFATWGLATCGICIFLGSSAFNALVNSAVLQQLSFLIPIALLLYOKRDPK 433
Db 358 KLOVPLNGLYLTMPCVIVFCIFLGSSSAFNATITSSSVVMDLVAYGIPAIACLRGR--Q 415
QY 434 FLPESTRAFVLPGRGIFLVNVLAVVFTSVTVTFSPFLTPVTAASTMYNTSAIIGVALALG 493
Db 416 MIPE-RAFVLPQALQGMVNIETALAYISLTITVLFPELPVTSNMNMYCAAFGIVLIIS 474
QY 494 VLNWVHARKHYOGPHLELDGRVGA-EFQVGP 525
Db 475 AFQIVDGRKHFTGP--RVDAEVLAVESYSGP 505

RESULT 2
Q92U50 PRELIMINARY; PRT; 517 AA.
ID Q92U50
AC Q92U50;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE PUTATIVE AMINO ACID PERMEASE.
GN AT2G01170.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RA *Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.*;
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL: AC006200; AAD14517.1;
DR InterPro: IPR002293; AA_rel_permease.1.
DR InterPro: IPR002027; Amino_acid_permease.
DR Pfam: PF00324; aa_permeases; 1.
SQ SEQUENCE 517 AA; 55359 MW; 0718A73ED50B116D CRC64;
```

Query Match 16.8%; Score 451; DB 10; Length 517;
Best Local Similarity 27.2%; Pred. No. 1.3e-25;
Matches 135; Conservative 104; Mismatches 217; Indels 40; Gaps 16;

```
QY 39 KKQFGTITIVSLAFVICSNWAGISGLQALLAGGPVTLTYLILSTLVYICIAFLAEL 98
Db 30 KRDLVSFNSFAISPSIISVLGTITTYTGLRFGGTVTLVYGMWFLAGSFTMCVGLSMAEI 89
QY 99 TSVYPTAGQYHFASILAPKSNRSISYV--CGLVSLLSWIAIGSSVTWIPAAQIPALI- 155
Db 90 CSSYPTSGGLYWSANLAGRPWAPLASNMWTGCKVVVYVCLQW-AVTASVDVFSLAQLIOIVL 148
QY 156 ----AAYSHYTSYSDSHVFLIYEGVALVLLFNLFALKRNPVWHEIG-----FGLTIALF 206
Db 149 LSTGGRNGGGYKGSDFVWIGIHGGILFIHALLNSLPISVLSFGQLAALNLLGLVLM 208
QY 207 VISFTAILARSNPKAPNSQVMTAWSNYTG---WSDGVCFFILGLSTSCFMFGLDAAMHLA 263
Db 209 LIPLV-----SPERATTKFVTFNFTDNGLGITSYAVIFVGLLMSOYTTITGYDASAHMT 263
QY 264 EECTDAARTVPKAVVSAAIIIGCTAFPTYTIAVLYGTDLSILS-----SAGYIPETWTQ 319
Db 264 EETVDADKNGPRGIIISAIGISILFGWGYILGISYAVTDIPSLSETNNSGGYIAIEFLYL 323
QY 320 SLRSL--SFATVLSGGI--VMAFALNAVOETASRLTWSFARDNGLVFSHTLERHPRW 375
Db 324 AFKNREGSGTGIVCLGVAVAVFVCGMSSTVSNSEMAFAESRDGAMPHSPLMHKVNRS- 382
QY 376 QVPVMSLFATW--GILATCGCI-FLGSSAFNALVNSAVLQQLSFLIPIALLLYOKRDP 432
Db 383 EVPI--NAVWLSALISFCMALTSLSIVAFQAMVSIATIGLYIAYAIPIILRVTLARN- 438
QY 433 KLPSTRAFVLPGRGIFLVNVLAVVFTSVTVTFSPFLTPVTAASTMYNTSAIIGVALAL 492
Db 439 TEVPG--PFSLGK-YGMVVGWAVLWVVTISVLSFSLPVAVPTTRETNLNTPVAVAGLVAI 495
QY 493 GVLNVHARKHYOGP 508
Db 496 TLSYWLFSARHWTGP 511

RESULT 3
Q22509 PRELIMINARY; PRT; 516 AA.
ID Q22509
AC Q22509;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE PUTATIVE AMINO ACID OR GABA PERMEASE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OX eurosids II; Brassicales; Brassicaceae; Arabidopsi.
RN NCBI_TaxID=3702;
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Turano F.J., Thakkar S.S.;
RA Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF019637; AAB71542.1;
DR InterPro: IPR002293; AA_rel_permease.1.
DR InterPro: IPR002027; Amino_acid_permease.
DR Pfam: PF00324; aa_permeases; 1.
SQ SEQUENCE 516 AA; 55331 MW; A4F88B0B4AFF8277 CRC64;
```

Query Match 16.7%; Score 449.5; DB 10; Length 516;
Best Local Similarity 26.7%; Pred. No. 1.7e-25;
Matches 132; Conservative 106; Mismatches 219; Indels 37; Gaps 14;

```
QY 39 KKQFGTITIVSLAFVICSNWAGISGLQALLAGGPVTLTYLILSTLVYICIAFLAEL 98
Db 30 KRDLVSFNSFAISPSIISVLGTITTYTGLRFGGTVTLVYGMWFLAGSFTMCVGLSMAEI 89
QY 99 TSVYPTAGQYHFASILAPKSNRSISYVCGVLSLWSIAIGSSVTWIPAAQIPALI--- 155
Db 90 CSSYPTSGGLYWSANLAGRPWAPLASNMWTGCKVVVYVCLQW-AVTASVDVFSLAQLIOIVL 148
```

Db 90 CSSYPTSGGLYYWSAMLAPRWAPLASWMTGWENIVGQWAVTASVDFSLAQIQLVILLS 149

QY 156 --AAYSHYTSQDSWHVELIYEGVALVLEFNLEALKRNPWVHEIG-----FGLTIALFVI 208

Db 150 TGRNCGGKYGSDFFVIGIHGILFIHALLNSLPSVLSVLSGQAAALNLLGLVLMILI 209

QY 209 SFIAIARNPAPNSQWWTANSVTG---MSDGVCFILGLSTSCFPMFIGIDAMHIAEE 265

Db 210 PLV-----STERATTKFVTNFNTDNGLGITSYAFIFVGLGLMSQYITITGYDAHNT 264

QY 266 CTDAAATVPKAVVSIIIGFCTAFPTTIAVLYGTTIDLDLSILS-----SAGVIPETMTQSL 321

Db 265 TVDADKNGRPGIIISAIGLSILFGWYILGISYAVTDIPSLSETNNSGGYAIALFVLA 324

QY 322 RSL--SFATVLSGGI--VMAPFALNAVOETASRLTWSFARDNGLVSTILRIHPRQV 377

Db 325 KNRFGSGTGGIYCLGVAVAVAFVFCGMSSTNSRMAYAFSRDGAHNSPLHVKVNSR-EV 383

QY 378 PVMSLFATW--GILATCGCI-PLGSSAFNALVNSAVVLOQLSLPIALLLYQKRPK 434

Db 384 PI--NAVLSALISFCMALTSGLSIVAFQAMVSIATIGLYIAIPIILRVTLARN-TF 439

QY 435 LPSFATVLSGGI--VMAPFALNAVOETASRLTWSFARDNGLVSTILRIHPRQV 494

Db 440 VFG--PFSLGK-YGMVGVAVLVVTVISVLSPLFVAVPIPAETLNTYTPVAVAGLVAIL 496

QY 495 LNMVHARKHYQGP 508

Db 497 SYWLESARHWTGP 510

RESULT 4

Q60113 PRELIMINARY: PRT; 542 AA.

ID O60113

AC O60113

DT 01-AUG-1998 (TREMBLrel. 07, Created)

DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE HYPOTHETICAL AMINO-ACID PERMEASE C15C4.04C.

GN SPBC15C4.04C.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RA Lyne M., Rajandream M.A., Barrell B.G., Xiang Z., Aves S.;

RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC -1- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.

DR EMBL: AL023290; CAA18895.1;

DR InterPro: IPR002293; AA_rel_permease.1.

DR pfam: PF00324; aa_permeases.1.

DR PROSITE: PS00218; AMINO-ACID_PERMEASE; 1.

KW Hypothetical protein; Transport; Amino acid transport; Transmembrane.

FT TRANSMEM 95 115

FT TRANSMEM 217 237

FT TRANSMEM 255 275

FT TRANSMEM 298 318

FT TRANSMEM 348 368

FT TRANSMEM 402 422

FT TRANSMEM 425 445

FT TRANSMEM 469 489

FT TRANSMEM 500 520

FT TRANSMEM 396

SQ DOMAIN 542 AA; 59726 MW; 17D9BI5C04299468 CRC64;

Query Match 16.5%; Score 445; DB 3; Length 542;
Best Local Similarity 29.3%; Pred. No. 3.8e-25;

Matches 146; Conservative 83; Mismatches 229; Indels 40; Gaps 16;

QY 36 PLEKKQFCFTITVSLAFVICSNWAGISGSLQALLAGGPVTLVGLISLVVVICIAFSL 95

Db 55 PVKREFSTWTFSEAFSISGLFATVTVTSYPLISGAPSAVWCWLIAGAGCMCIASLV 114

QY 96 AELTSSVYTAGQVHFASILAPKSNRSISYVCGVLVLSWIAIGSVTWMIPAAQIPALI 155

Db 115 AELVSAYPTSGLYFTCKDLVPAKSMFVAVWVGLNLGQAAGVSTDWSCAQ---LLL 171

QY 156 AAYSHT---YSDSWHVLIYEGVALVLEFNLEALKRNPWVHEI-GFGLT--IALFV 207

Db 172 AAVSISTDLKYPINQHIV---GVMAAVIVFHGLVNSLSTRWLDRTREYATFHLLV 227

QY 208 ISFIAIARNPAPNSQWWTANSVTG---MSDGVCFILGLSTSCFPMFIGIDAMHIAEE 266

Db 228 VCMICLLAKCPKFTGKYVETDQASGCHWHPDGFSLFGLSVANWCMTDYDATAHIAEEI 287

QY 267 TDAARTVPKAVVSIIIGFCTAFPTTIAVLYGTTIDLDLSILSSAGYIP---PETMTQSL 321

Db 288 ENAAVRAPNAIALALSITYVLGVNENIVLAFTMCTDLDSLNSLSELGQVPAQIEYNVLGKK 347

QY 322 RSLFATVLSGGI--VMAPFALNAVOETASRLTWSFARDNGLVSTILRIHPRQV 381

Db 348 GSAF-TLSF--IINFTGITAMQANA-RTIWAFSRDQALPFSRVYWKINKTTTPV-- 401

QY 382 LPATWGILATC---GCIFLGSSTAFNALVNSAVVLOQLSLPIALLLYQKRPKLPST 438

Db 402 -LAVLVNVFVFCIALNLIGLSIEAIEAFSCAIALDWSYVPIACKLFGKRLNPKG- 459

QY 439 RAFVLPRGIGF---LVNVLAUVFTSVTVFSPFLTVPTAASTMNTSAIIGVALAGVL 495

Db 460 -----PWLGNWASHEFNAYAVCVTAFTSVIFLMPETVRPTQNMNNAVAVVLAGVLSLV 514

QY 496 NMVHARKHYQGHLELD 513

Db 515 YWMSGARKSYIGPRINVD 532

RESULT 5

Q94CQ2 PRELIMINARY: PRT; 556 AA.

AC Q94CQ2

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DE P0660F12.28 PROTEIN.

GN P0660F12.28

OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzeae; Oryza.

OX NCBI_TaxID=4530;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Cv. NIPPONBARE;

RA Sasaki T., Matsumoto T., Yamamoto K.;

RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC clone:P0660F12.28";

RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AP003687; BAB3862.1;

SQ SEQUENCE 556 AA; 59916 MW; 4B13A088F303A197 CRC64;

Query Match 16.1%; Score 432.5; DB 10; Length 556;
Best Local Similarity 28.2%; Pred. No. 3.3e-24;
Matches 142; Conservative 92; Mismatches 213; Indels 57; Gaps 16;

QY 39 KQFGTITVSLAFVICSNWAGISGSLQALLAGGPVTLVGLISLVVVICIAFSLAE 98

Db 76 KRGLSVNSNFAFSFSLISLVGTVTTTGTGLRGYGPVSMTLGLVVAFAAFNCVALSMAEI 135

QY 99 TSVYPTAGQVHFASILAPKSNRSISYVCGVLVLSWIAIGSVTWMIPAAQIPALI--- 155

```
Db 136 CSATPTSGGLYYSKLAGKEMAPLASWVTGWFNMGOWACTTSVDFSLAQLQVILLA 195
Qy 156 --AAYSHYTSQDSWHVFLIEGVALVLLFNLFALKRNPWVHEIGFGLTIALFVIFSTAI 213
Db 196 TGGANGGGMASKYVVLAIYSAILLHGLINSPLRWLSWFGQLG-----APWNV 245
Qy 214 LARSNKAPNSQVWTAWNS--NYTGWSDGVCFIL--CLSTSCFMFGLDAAHLEBECTDA 269
Db 246 AAVAKERSPEEFITFNAENGAGIHD--KAYILAGLLMSQVSLGYDTSAHIEETKNA 304
Qy 270 ARTYKAVSAIIIGFCTAPPTIAYLYGITDLSILS-----SAGYIPFETMTQSL--RS 323
Db 305 DWSGPIGITSVALSTMFNGMIYLIALTSTMTDIPYLLNPSNDAGGAYIAQALYTSFHRRY 364
Qy 324 LSPATVLSGGI--VMAFFALNAVOETASRLTWSFARDNGLVFSTHLRIHPKQVP--- 378
Db 365 GTGAGALACLGVIATAIFCGSACITSNRMGYAFSRDGMPLSRVHWRVDSR--EVPLNV 423
Qy 379 VWSLFATWGLATCGCIGFSSAFNALVNSAVVLOQLSLPIALLLYQKRDKFLPST 438
Db 424 VMLSVAFAVFWALTS---LGSQVAFQAMVSIAT---LGLLIACAL-----PVFERVT 469
Qy 439 --BAFLVPRG-----IGFLVNLAVVFTSVTVFFSPFLVPTAASMTYSAIIGVA 489
Db 470 TARSAFV--KGPHLGKYGIVGWGVVWVATVTLFSLPVAIPYPAKETFNFTPVAVGGV 527
Qy 490 LALGVNLNVVHARKHQQPHLELD 513
Db 528 LLLSLVAVWLHARFVQGPVTNVD 551

RESULT 6
Q94CQ3 ID Q94CQ3 PRELIMINARY; PRT; 515 AA.
AC Q94CQ3;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE P0660F12.27 PROTEIN.
GN P0660F12.27
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxId=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0660F12.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003687; BAB63861.1; -.
SQ SEQUENCE 515 AA; 55279 MW; 97BEF49648BA32A1 CRC64;

Query Match 15.2%; Score 410.5; DB 10; Length 515;
Best Local Similarity 25.8%; Pred. No. 1.3e-22;
Matches 133; Conservative 92; Mismatches 219; Indels 71; Gaps 14;

Qy 39 KQPGTTIVSLAPVTCNSWAGISGLQALLAGGPVTLGYLILSTLVYICIAFSLAE 98
Db 31 KRGLSLNSNFAFSFISIVLTGTYTGLRYGVPVSWTLGWLVSFAFNCVALSNAEI 90
Qy 99 TSVYPTAGGYHFAFSLAPKSNRSISYVCGLVSLISWIAIGSSVTMIPAOQIPALI--- 155
Db 91 CSAYPTSGGLYYSKLAGKEMAPLASWVTGWFNIVGOWACTTSVDFSLAQLQVILLA 150
Qy 156 --AAYSHYTSQDSWHVFLIEGVALVLLFNLFALKRNPWVHEIGFGLTIALFVIFSTAI 213
Db 151 TGGANGGGMASKYVVLAIYSAILLHGLINSPLRWLSWFGQLG----- 195
```

```
Qy 214 LARSNKAPNSQVWTAWNSYTC--WSDGVCFILGLSTSCFMFGLDAAHLEBECTDAART 272
Db 196 -----APWNVANGMGIHQKAYILAVGLLMSQVSYGIDYDTSAHIEETKNAIDS 244
Qy 273 VPKAVYSAILIIGFCTAPPTIAYLYGITDLSILS-----SAGYIPFETMTQSL--RSLSF 326
Db 245 GPMGIITSVGLSTFMFCWIIYLIALTSTMTDIPYLLNPSNDAGGAYIAQALYTSFHRRYGTG 304
Qy 327 ATVLSGGI--VMAFPALNAVOETASRLTWSFARDNGLVFSTHLRIHPKQVP---VWS 381
Db 305 AGALACLGVIATAIFCGSACITSNRMGYAFSRDGMPLSRVHWRVDSR--EVPLNVVWL 363
Qy 382 LFATWGLA-----TCG--CIF-----LGSSTAFNALVNSAVVLOQLSLF 418
Db 364 SVAVAFVMAVTNSKISIAAAAVNFSSCDLVVFSQSLGSOVAFQAMVSVTTLGLYIAY 423
Qy 419 LPIALLLYQKRDKFLPSTRAFLVPRGIFLVNLVAVFTSVTVFFSPFLVPTAAS 478
Db 424 ALPVFRVTTAR--KSEVPG--QFHLGR--YGLMVGMMVWVWVATVTLFSLPVAIPYPAKET 479
Qy 479 MNYTSAIIIGVALGVLNVVHARKHQQPHLELD 513
Db 480 FNTVPVAVGGVLLSLVAVWHAREFWFQGPVTNVD 514

RESULT 7
Q94CQ6 ID Q94CQ6 PRELIMINARY; PRT; 521 AA.
AC Q94CQ6;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE P0660F12.24 PROTEIN.
GN P0660F12.24
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxId=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0660F12.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003687; BAB63858.1; -.
SQ SEQUENCE 521 AA; 55915 MW; 858AB5BBF5EDA2C8 CRC64;

Query Match 14.6%; Score 394; DB 10; Length 521;
Best Local Similarity 26.4%; Pred. No. 2.2e-21;
Matches 133; Conservative 96; Mismatches 230; Indels 44; Gaps 18;

Qy 39 KQPGTTIVSLAPVTCNSWAGISGLQALLAGGPVTLGYLILSTLVYICIAFSLAE 98
Db 32 KRGLSALSNAFSAFISIVMGVTTYNTGLRYGVPVSWTLGWLVSFAFNCVALSMAEI 91
Qy 99 TSVYPTAGGYHFAFSLAPKSNRSISYVCGLVSL--SWIAIGSSVTMIPAOQIPALI-- 155
Db 92 CSAYPTSGGLYYSKLAGKQWASLASWVTGWFNVVGOWAAI--ASVDFSLAQLQVILL 150
Qy 156 ---AAYSHYTSQDSWHVFLIEGVALVLLFNLFALKRNPWVHEIGFGLTI-----ALFVI 208
Db 151 STGGNGGGYMAKSYTLVLIY---AFILHLGILNSLPIEWSLFGHVATWNAAGFVL 207
Qy 209 S-FIATLARNKAPNSQVWTAWNSYTCNSWGVCFI--LGLSTSCFMFGLDAAHLEEC 266
Db 208 TILIPAVAKDRNIEFVTHLNTENGMIHDKAYILAVGLLMSQVSYGIDYDTSAHVEET 267
Qy 267 TDAARTPVKAVYSAILIIGFCTAPPTIAYLYGITDLSILS-----SAGYIPFETMTQSLR 322
Db 268 KNADRSGPITGIITSVLFATVPGWIVLALTSVVTDIPVLLSPSNDAGGY-----AIAQALY 323
```


QY 323 SL-----SPATVLSGGIVMARFALNAVQETA-SRLTWSFARDNGLVFSTHLERHP 374
 Db 324 TAFHRRYSGVGGIVCLGAVAVAVFLGACVTSNSRWAYFSRGGAMPLSRVYRN-K 382
 QY 375 WQVP---VMSLFATWGILATCGCIFLGSSTAFNALVNSAVVLOOLSLFIPALILYOKRD 431
 Db 383 HEVPLNVNVLGVAVAFVMAITS---LGSQVAFQAMGSIATLGMVIAVALPVEFRVTTAR- 438
 QY 432 PKFLSTRAFLVPRGIGFLVNLVAVFTSVTTVPFSPPLTVPTA-ASTMNTYSIIIGVAL 490
 Db 439 RSFVPG--PFHLGR-YGVVGVWAGVAVVATVTLFSLPVPVANKETFTNPVAVGGVL 495
 QY 491 ALGVNLVNVVHARKHYQGHLELD 513
 Db 496 LLSVGAWLRLARFVQGPITNV 518
 RESULT 8
 Q94CQ5 PRELIMINARY; PRT; 532 AA.
 AC Q94CQ5;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE P0660F12.25 PROTEIN.
 GN P0660F12.25.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 RT clone:p0660f12.25";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF003687; BAB3859.1; -.
 SQ SEQUENCE 532 AA; 56603 MW; 6A427F02A2DB87C8 CRC64;

Query Match 14.5%; Score 389; DB 10; Length 532;
 Best Local Similarity 26.2%; Pred. No. 5,3e-21;
 Matches 136; Conservative 96; Mismatches 217; Indels 70; Gaps 20;
 QY 39 KKGFGTITVSLAFVICSNWAGISGSLQALLAGGPVTLGLISTLVYICIAFSLAE 98
 Db 26 KRLGLSTLSNFAFSFANISVLTGTATYNTGLRYGGVPSMTLGLVYALFNGCVALSMAEI 85
 QY 99 TSVPYTAGGYHFASITLAPKSNRSTSYVGLVSL-SWIAIGSSVTWIPAOQIPALI-- 155
 Db 86 CSAYPTSGGLYHSAKLACKEWASLASWTGHNIVGOWAI-ASVDFSLAQLLOVILL 144
 QY 156 ---AAYSHYSQDSWHVFLIYEGVALVLLFNLFALKRNWVHETG-----FGLTALFV 207
 Db 145 STGGANGGGYMASKYVVLVI---CAVILIHGVINSPLQWLSFGQICAIWNAAGVFL 201
 QY 208 ISFIATLARNKAPNSQVTAWSNTGW---SDGVCFILGLSTSCFMEFIGDAAMHAE 264
 Db 202 VILIPAVADRPVS--EFVTHLNTDNGMGIHSAKAYIALVGLMSQYSLVGDYDSAHWE 259
 QY 265 ECTDAARTVPKVASIIIGCTAFPTIYAVLIGITDLSLS-----SAGYIPFETMQS 320
 Db 260 ETKNADRSFGPIITSWCATVFGWIYLLALTAVTDIPYLLSPGNDAGY----AIAQA 315
 QY 321 L-----RSLSFATVLS-CGGIVMAFFALNAVQETASRLTWSFARDNGLVFSTHLERHP 373
 Db 316 LYTAFHRRYSGAGGTACLGAVAVAFVMAITS---LGSQVAFQAMGSIATLGMVIAVALPVEFRVTTAR- 368
 QY 374 RWQVP---VM-----SLFATWGILATCGCIF-----LGSSTAFNALVNSA 410

Db 369 KHEVPLNVNVLAVAVAFFMALTVNYIPSTCRCCFCSSCVRCSDFSLGSQLVAFQALGSTA 428
 QY 411 VVLOOLSELIPALILYOKRDKPFLPSTRAFLVPRGIGFLVNLVAVVFTSVTTVPFSPRL 470
 Db 429 TIGMYTAYLAPVEFRVTTAR-RSFVPG--PFNLGK-YGVLVGVWGVVAVVTVLFSLPV 484
 QY 471 TVPTA-ASTMNTYSIIIGVALGVLNVLNVVHARKHYQGP 508
 Db 485 AYPVANKETFTNPVAVGGVLLLSVGAWLHAREFWQGP 523
 RESULT 9
 Q9KZF1 PRELIMINARY; PRT; 504 AA.
 AC Q9KZF1;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PROBABLE AMINO ACID/METABOLITE PERMEASE.
 GN SC8F11.28C.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Seeger K.J., Harris D.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Thomson N.R., parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
 RA Kinashl H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb streptomycetes coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL; AL353864; CAB88979.1; -.
 DR InterPro; IPR002293; AA_rel_permease.1.
 DR InterPro; IPR002027; Amino.acid.permease.
 DR PROSITE; PS00218; AMINO_ACID_PERMEASE; UNKNOWN_1.
 SQ SEQUENCE 504 AA; 53812 MW; 5B45C4950F0E055A CRC64;

Query Match 14.4%; Score 387; DB 2; Length 504;
 Best Local Similarity 25.7%; Pred. No. 7.1e-21;
 Matches 127; Conservative 92; Mismatches 223; Indels 52; Gaps 13;
 QY 36 PLEKKQPGTITVSLAFVICSNWAGISGSLQALLAGGPVTLGLISTLVYICIAFSL 95
 Db 23 PVLARRMGCGFNFAISFVSISILSGCMTLYGCMGTGGPAMVLMWAGVGLFVLCVGMAL 82
 QY 96 AELTSVYPTAGGYHFASITLAPKSNRSTSY--VCLGVLSLLSWIAIGSSVTWIPAOQIPA 153
 Db 83 AEVTSAYPTSGALYVMAADRLGRRGWYTCWMLNLLGLGAIGIDYGA----- 130
 QY 154 LIAAYSHYSQDSW-----HVFLIYEGVALVLLFNLFALKR-----NPWHVEIGF 199
 Db 131 --ALFTGAPINLQMGFEPTPEKTMFLFVLLHATLNLFGVRLVSLNSISVWHLAV 188
 QY 200 GLTALFVIFIAFLARSNKAPNSQVTAWSNTGWSDGV-CFILGLSTSCFMEFIGLDA 258
 Db 189 ALIVGALVI-----VPDHHQSP-SFVTFEYVNDTGWENPLYVAAIGLLAQYTSGYDA 241
 QY 259 AMHIAEECTDAARTVPKVASIIIGCTAFPTIYAVLIGITDLSLSA-GYIPFETM 317
 Db 242 SAHLSETSNASVSAARGIVRAIWSVWLAGVLLGTAIQLQDYDATROTATGVPVPAQIL 301

Db 369 VAAITWATASILSVLVWG-----STLVSVAGTSAYTIIVVSVCTVIFLFSFTVPIVGLWLWG 424
 Qy 431 DPKEPSTRAFLPRGIFLVNVLAVVFTSVTVFFSFPLTVPTAASTMYTSIIIGVA- 489
 Db 425 TPKW-DKMGPNMGKGVFMLFAVLSIV--SMILIF-----VIGIQP 462
 Qy 490 -----LALGVNLVNVHARKHYQGFHLELGRVUGA 519
 Db 463 PNDWALYITVGFILTAIVWFAFERNRFQGPPL---GDIIAA 501

RESULT 12
 Q92NI8 PRELIMINARY; PRT; 515 AA.
 AC Q92NI8
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE PUTATIVE AMINO-ACID PERMEASE PROTEIN.
 GN SMC01597.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21368234; PubMed=11474104;
 RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
 RA Barloy-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G.,
 RA Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
 RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
 RA Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gurjal M.,
 RA Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
 RA Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
 RA Masuy D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,
 RA Ramsperger U., Surzycki R., Thebaud P., Vandenbol M.,
 RA Vorholter F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
 RP "The composite genome of the legume symbiont Sinorhizobium meliloti.";
 RL Science 293:668-674(2001).
 DR EMBL; AL591790; CA646793.1; -.
 KW Complete proteome.
 SQ SEQUENCE 515 AA; 55915 MW; 956BD9C62060A524 CRC64;

Query Match 11.4%; Score 307.5; DB 16; Length 515;
 Best Local Similarity 22.7%; Pred. No. 5.8e-15;
 Matches 120; Conservative 99; Mismatches 230; Indels 79; Gaps 19;

Qy 37 LEKK--QFGTTIVSLAF-VICNSWAGISGSLQALLAGGPTVLLYGILISTLVYICIAF 93
 Db 25 LERRMSQSFNAV---SFSIICILSGGINSLAQATSGAGG-AAIGIGWPLGCFISFVFAV 80
 Qy 94 SLAEITSVYPTAGGYHFASILAPKSNRSISYVCCVLSLLSWIAIGSSVTMIP 147
 Db 81 AMAQISSAYPTAGGLYHWSILG---NRFTGLTAWFNLLGLVTLGALNVGTVYFFMG 136
 Qy 148 AQQIPALIAAHSHTYSQDSWHVFLIYEGVALVLLFNLFALKRNPWWHEIGFGLT----- 202
 Db 137 SFGTP-----YFGLDDTTMTTRIVFLAIIITGAQALVNHMGIGLTAKLTD 179
 Qy 203 -----IALFVISFIATLARSNKPAPNSOVWTAWSNYTCWSDG-----VCFILGL 246
 Db 180 FSGYLIFVTAIIALLVCLAAADTVEI--SRFLT-PSNYSETGNVWPQTSQGWVFLGL 236
 Qy 247 STSCPMFICGLDAAMHLEAECTDAARTVPKAVWSAIIIGCTAFPTIYAVLIGITDLSIL 306
 Db 237 LLPIYITGYDASAHTSBETVKAHSPVPRGMISSVLWAAALFGVIMLCSFVLMIPDMDAA 296
 Qy 307 SSAGYIPETWTQSLRSLSFATVLSGGIYNWAFALNAVQETASRLTWSFARDNGLVFT 366
 Db 297 KOGNVVFWAMDQOVNPI-IKDVLVLAIFVSQWLCGLIATVTSVSRMIFAFSRDGGLPASK 355

Qy 367 HLERHPRMOPVPMVNSLFTATWGLIAT-----CGCIFLGSSTAFNALNSAVVLOQLSFLIP 421
 Db 356 ALSXSVSTYRTPVAAIW-TGSILAVLVWGSLSVSGIDTPVTIVVSVCTVIFLFFSFAIP 414
 Qy 422 IALLYKROPKFLPSTPRAFLPRGIFLVNVLAVVFTSVTVFFSFPLTVPTAASTMY 481
 Db 415 ITLGLFWANGTSKW-DKMGPNNLGEGMKFLAIVSIV-----AMVLIFVLGIQPPNDWALY 468
 Qy 482 TSAIIGVALGVLNVVHARKHYQGFHLELGRVUGA 519
 Db 469 IT--VGLIIVTGII-WFAFENRRFRGPPIGDEVARRQAEIAAAEKAYG 513

RESULT 13
 Q9C0Z0 PRELIMINARY; PRT; 543 AA.
 AC Q9C0Z0
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE PUTATIVE AMINO-ACID PERMEASE.
 GN SPAPB24D3.02C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972H-;
 RA Collis M., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL590582; CAC36898.1; -.
 DR InterPro; IPR002293; AA_rel_permease_1.
 DR InterPro; IPR002027; Amino-acid_permease.
 DR Pfam; PF00324; aa_permeases; 1.
 SQ SEQUENCE 543 AA; 59257 MW; 1B414DA8EAA9E762 CRC64;

Query Match 11.3%; Score 303; DB 3; Length 543;
 Best Local Similarity 24.1%; Pred. No. 1.3e-14;
 Matches 123; Conservative 109; Mismatches 227; Indels 52; Gaps 19;

Qy 39 KQFGTTIVSLAF---VICNSWAGISGSLQALLAGGPTVLLYGILISTLVYICIAFSL 95
 Db 44 KREFSLLAVFGQSGSGMGLCPSLVC-SNAFSGMNCAGG---MVMSWFGVATCLLPALAFAL 99
 Qy 96 AELTSVYPTAGGYHFASILAPKSNRSISYVCCVLSLLSWIAIGSSVTMIPAQQIPALI 155
 Db 100 SELASSMPTSGSLVFTAYLSPPKYRAPLWFLGVLVLAIAITGTFASTIYAAAGLVQATA 159
 Qy 156 AAYSHYSQDSWHVFLIYEGV-----ALVVL-----LNFALKRNPWWHEIGFGLTIA 204
 Db 160 SVANPSYAPTAYEEYGVIVVALSFACSAIIVLPKFLARFSFNV-----VFQI---CTIL 211
 Qy 205 LFVTSFIALLARSNKPAPNSOVWTAWSNYTCWSG-VGCFILGLSTSCFMPIGIDAAHMLA 263
 Db 212 IFIIS-LAASSTSTRNTGSIYFNGFNYSQWNTMGSFILCFTTPVWVLSGSCATV 270
 Qy 264 EECDAARTVPKAVWSAIIII---GFCATPPTIYAVLIGITDLSLSSAGYIPFETMTQ 319
 Db 271 EEAKNASKAAPAIATISSLTVSLFMGFCIMI--TIAGTMG-HDFSSILNTYGPVPSOVLY 327
 Qy 320 SLRSLSFATVLSGGIYNWAFALNAVQETASRLTWSFARDNGLVFTHLERHPRMOPV 379
 Db 338 NNLGKRGAVGSVAVLIILALCFNCALCLASSREIFAFAKDKGLPGSMIFKKLTP-GGILP 386
 Qy 380 WSLFATWGLATCGCIFLGSSTAFNALNSAVVLOQLSFLIP-IALILYKRDIP-KFLPS 437
 Db 387 NAILLVNLYTIIVGLLNVVNTALISSIFNLIAITAFFISYLSPLVCLLFLNRLNPGKFY-- 444
 Qy 438 TRAFVLPRGIFLVNVLAVVFTSVTVFFSFPLTVPTAASTMYTSNAIIGVALGVLNW 497

```
Db 445 CCKFSKP-----ISIVAVANLWFMALMLFPSPVONPKVEMNNAIVVLGFTFFFCVGY 498
Qy 498 VV---HARKHYGPHLELDGRV---VGAEFQ 522
Db 499 YLPKYGGKTFKPKVPKVTDENVTGVTWDFQ 529

RESULT 14
Q9RKN0 PRELIMINARY; PRT; 389 AA.
AC Q9RKN0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE POSSIBLE AMINO ACID/METABOLITE PERMEASE (FRAGMENT).
GN SCD17.01.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Brown S.P., Harris D.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL118515; CAB56378.1;
DR InterPro; IPR002293; AA_rel_permease_1.
DR InterPro; IPR002027; Amino_acid_permease.
DR Pfam; PF00324; aa_permeases; 1.
FT NON_TER 1
SQ SEQUENCE 389 AA; 41731 MW; FCBE62374516AE3F CRC64;

Query Match 10.58; Score 284; DB 2; Length 389;
Best Local Similarity 26.18; Pred. No. 2.4e-13;
Matches 97; Conservative 59; Mismatches 150; Indels 66; Gaps 13;

Qy 171 LIYEGVALVLLNFALK-----RNPVNHGIEGGLTIALFVIFAILARSNPKAPN 223
Db 35 LIFAAILLLHGLLNTFGVRIVGLLNSVVMHVLG-----VAIVGALTFAPD 82

Qy 224 -----SQVWTASVNTGSGVCEIL-GLSTSCFMFIGLDAAMHLAECTDAARTVPKAV 277
Db 83 HHQSASFVGFVNNTGSGVYVLLIGLLMAQYTFGYDASAHMTEETHDASTAGPKGI 142

Qy 278 VSAILIGFCTAPYTIYVLXIGTDLDSLS-ACYIPFETMTQSLRSL-----FAT 328
Db 143 VRSIWTSMIAGFVLLGFTFAIOSYDGLTSPGAPPAQTLDDALGATAGKLLLVVIGA 202

Qy 329 VLSCGGIVMAFFALNAVQETASRLTWFSFARDNGLVSTHLERHPWOPVWVSLFATWGI 388
Db 203 QLFCG---MASVTAN-----SRMIYAFSRDGLAPYSHIWHSVNRTPTV---AAVW-- 248

Qy 389 LATOCIFLG-----SSTAFNALVNSAVLQQLSFLIPIALLYQ-----KRDPKFLPSTR 439
Db 249 LAALAAVLGLPYLINVTAAVTSIAVIGLYIAYVIPTLLRVKGAAPFSGPWHLGRWS 308

Qy 440 AFVLPGRGIGLVNVLAVVFTSVTTFVFPFLTVPTAASTNNYTSIAIGVALALGVLNWWV 499
```

```
Db 309 Q-----LVGVAVTWVGVIITVLFMLPQVSPVTTWETTFNADIAVLAVLGFATWMLV 359
Qy 500 HARKHYGPHLE 511
Db 360 SARHWFNLNPDHE 371

RESULT 15
Q9LLJ2 PRELIMINARY; PRT; 476 AA.
AC Q9LLJ2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PUTATIVE AMINO ACID TRANSPORTER.
GN SCC46.09C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Brown S.P., Harris D.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL118515; CAB575313.1;
DR InterPro; IPR002293; AA_rel_permease_1.
DR InterPro; IPR002293; Amino_acid_permease.
SQ SEQUENCE 476 AA; 49188 MW; H3DB5592BC3FC0C3 CRC64;

Query Match 9.38; Score 250.5; DB 2; Length 476;
Best Local Similarity 23.58; Pred. No. 9.1e-11;
Matches 125; Conservative 85; Mismatches 209; Indels 113; Gaps 22;

Qy 17 TTKNTETAAAGCAGESLNVPLE---KKQFGT-----ITIVSLAFVICSNWAGIS 62
Db 4 TLRPAETAAPVPPDTPAPAPSESTLKRISGVVGGTLLTLSCVTPASTLEVVV---PDLF 60

Qy 63 GSLQALLAGGPTVLLYGLISTLVYICIAFLAELTSTVPTAGGQVHFASILAPKSINR 122
Db 61 GSL-----CTATPAL-TIAGVSLCVAVAFCYSELOGLVPSAGGEVAMVSTMA---GR 108

Qy 123 SISVCGVLVSLLSWIAIGSSVTNIPAOQIIPALIAVSHYTSQDSWHV--FLIIEGVALV 180
Db 109 LAGLVFLVLSLL-----VYMI-----VPPVIAMGTADYLEPVLHLDPSLAGAGVMLA 156

Qy 181 LLENLFALKRNPVHVEIGFGLTIALFVIFAILARSN-PKAPNSQVWTASVNTGMSDG 239
Db 157 TLAGLLDLRANAWI--TGVFLVLEVIAAAVAVVGLFAHAERPGCSLVSMVEVGTTGGADP 214

Qy 240 VCFIL---GLSTSCFMFIGLDAAMHLAECTDAARTVPKAVNSAIIIGFCTAPYTIYVL 296
Db 215 VTAMLVWGLAIAIFITQGFSTAVLSEELNPRNRVARTVLTATLAISTVILVPPVVAIT 274

Qy 297 YGITDLDSI-----LSSAGYIPFETMTQSLRSLFATVLSGGIVMAFFALNAVQETASRL 352
Db 275 FGASDLAELTGCDISAMVTAWSNSAVGTFFVSLCVALAIVNAGIVM-----VIQN--SRV 326
```

‘Mon’ Apr 29 09:00:46 2002

Qy	353	TWSFARDNGLVSTHLRIIPRQVDPV	-----WSLFTWGTGIIATCGCIIFTGS	399
Db	327	LFASARAKA	-----WPAENGLLGRGRGSPWATLAVGVPCAAALC	372
Qy	400	STAFNALVNSAVLQQLSFLLIPIALLLYOKRKDPLPSTRAFLVPRGIGFLVNVLAVFT	459	
		: : : : : : : : : : : : : : : : : : : :		
Db	373	DTLYGV-----TGVSTMYLLVAVAAALLSRGHAHHAHWRMPLVPVVFYLLIATVAYILT	429	
		: : : : : : : : : : : : : : : : : : : :		
Qy	460	SVTTVFSSFLVPTAAASTMNYTTSALIGVALGALGVNVVHARKHQGPHLE	511	
		: : : : : : : : : : : : : : : : : : : :		
Db	430	Q-----QEASHLLTGSTAAVA-----TLYWALYLR-----PRLD	459	

Search completed: April 27, 2002, 07:57:47
Job time: 703 sec



Result No.	Query	Score	Match	Length	DB	ID	Description
1	681.5	25.3	563	1	HNML_YEAST	P19807	saccharomyce
2	540.5	20.1	561	1	BI05_YEAST	P53744	saccharomyce
3	454	16.9	567	1	YF1L_SCHPO	Q9us40	schizosacch
4	429	15.3	557	1	YC9D_SCHPO	Q74537	schizosacch
5	420	15.6	544	1	IC9D_SCHPO	Q09887	schizosacch
6	409.5	15.2	551	1	AAP2_NEUCR	O59942	neurospora
7	383.5	14.2	571	1	UGAA_YEAST	P32837	saccharomyce
8	351.5	13.1	550	1	YA08_SCHPO	Q10087	schizosacch
9	334	12.4	553	1	GP7L_CANAL	Q74248	candida alb
10	321	11.9	618	1	YKRA_YEAST	P36029	saccharomyce
11	311.5	11.6	594	1	YF3L_SCHPO	Q59813	schizosacch
12	274	10.2	551	1	YFVJ_SCHPO	Q9ut18	schizosacch
13	225.5	8.4	461	1	YCFJ_ECOLI	P76037	escherichia
14	215.5	8.0	452	1	YEFJ_ECOLI	P33016	escherichia
15	199.5	7.4	440	1	YJ99_MYCTU	Q10858	mycobacteri
16	197.5	7.3	596	1	AG22_YEAST	P38090	saccharomyce
17	188.5	7.0	608	1	DIF5_YEAST	P53388	saccharomyce
18	181	6.7	636	1	CTR4_HUMAN	O43246	homo sapien
19	177.5	6.6	602	1	GAP1_YEAST	P19145	saccharomyce
20	177.5	6.6	657	1	CTR2_MOUSE	P18581	mus musculus
21	177	6.6	499	1	ANSP_ECOLI	P77610	escherichia
22	175	6.5	558	1	AGP3_YEAST	P43548	saccharomyce
23	174.5	6.5	465	1	Y093_RHIME	O87394	rhizobium m
24	174.5	6.5	658	1	CTR2_HUMAN	P25269	homo sapien
25	170.5	6.3	418	1	YJBE_ECOLI	P39277	escherichia
26	170	6.3	469	1	YGBP_BACSU	P46349	bacillus su
27	169.5	6.3	574	1	YHC6_SCHPO	O60170	schizosacch
28	169	6.3	497	1	ANSP_SALTY	P40812	salmonella
29	168.5	6.3	603	1	HIP1_YEAST	P06775	saccharomyce
30	168	6.2	481	1	YJ79_MYCTU	Q10875	mycobacteri
31	167.5	6.2	445	1	YHPM_ECOLI	P45539	escherichia
32	165	6.1	624	1	CTR1_RAT	P30823	rattus norv
33	164.5	6.1	888	1	LYSP_ECOLI	P25737	escherichia

FT	TRANSEM	294	314	POTENTIAL.
FT	TRANSEM	343	363	POTENTIAL.
FT	TRANSEM	399	417	POTENTIAL.
FT	TRANSEM	427	445	POTENTIAL.
FT	TRANSEM	466	486	POTENTIAL.
FT	TRANSEM	492	512	POTENTIAL.
FT	CARBOHYD	7	7	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	20	20	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	248	248	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	341	341	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	563 AA;	62056 MW;	1A7E1D146DB7807 CRC64;

Query Match	25.3%;	Score 681.5;	DB 1;	Length 563;
Best Local Similarity	33.2%;	Pred. No. 9.7e-37;		
Matches 156;	Conservative 95;	Mismatches 210;	Indels 9;	Gaps

QY	39	KKQFGTTIVSLAPVICSNWAGISGSLQALLAGGVTLGLYLGLISTLVYICIAFSLAEL 98
DB	55	RKSFSLWSILCVGFGLTNSWFGISTSWAGISGGPMVYIIHIALISICIGTSLGEL 114
QY	99	TSVYPTAGGYHFASILAPKSNIRSIYVGLVSLLSWIAIGSSVTMPAQQIPALIAAY 156
DB	115	SSAYPHAGGQFWNSKLAPPKYKRFAYMCGSFAYAGSVFTSASTLSVATEVGVYALT 174
QY	159	SHNYSODSWHVELIYECVALWLLFNLFALKRNPWTHETGFGTIALFVISPFIILARSN 218
DB	175	HPFIPKRWIHFVCFELHLFLFMFFNCYG-KSLPIISSSSLYISLSFFETITTVLACSH 233
QY	219	PKAPNSD-VMTANSNTGWSD-GVCFITGLISTSCFVIGLDAAMHAECTDAARTVPKA 276
DB	234	GKFNDAKEVFATPNETGMKNGGIAFVGLINPAWFSCLDCDATHAFVEKPERVPIA 293
QY	277	VSAIIIGCTAPPYTIAVLYGTDLSDLSSSAGYIP-FETMTQSURSLSFATVLSGGGI 335
DB	294	IMGTVAIGVFTFCYVYIAAFSFIQDLDAVLSTGTAPILDIVNOALGNKSGAIFLGLCLII 353
QY	336	VMAFFALNAVOEASRLTWSFARDNGLVSTHLERHPRWQVPPVNSLPATWGILATCGGI 395
DB	354	FTSFGCVIACHTWOARLCWSFARDNGLPSURLWSQVNPHTGVPNLNAHLMSCAWITLIGLL 413
QY	396	FLGSSTAFNALNWSAVVLOQLSFLPIALLLYQKRDPKPLPSTRAFLPRGIGFLVNVLLA 455
DB	414	YLASSTAFOSLITGCI AFLLLSYIIPVICLLAKKRNIAGHP---FWLGR-KGFPSTNIVL 468
QY	456	VVFTSVTTTFFESFPLTPVPTAASTMYNTSAIIGVALGVLNVNVVHARKHY 505
DB	469	LGNVTSVYFSPFPVLPVTOKNNMVCVVIVGYTAYSILYKWKYAKGKEF 518

RECORD#	2
BIO5_YEAST	
ID	BIO5_YEAST STANDARD; PRT; 561 AA.
AC	P53744;
DC	DT 01-OCT-1996 (Rel. 34, Created)
DF	01-OCT-1996 (Rel. 34, Last sequence update)
DE	01-NOV-1997 (Rel. 35, Last annotation update)
DE	Transmembrane regulator of KAPA/DAPA transport.
GN	BIO5 OR YNR056C OR N3502.
OS	Saccharomyces cerevisiae (Baker's yeast).
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX	NCBI_TaxID=4932;
RN	[1]
RC	SEQUENCE FROM N.A.
RA	STRAIN=FL100;
RC	Phalip V., Jeltsch J.M., Lemoine Y.;
RL	Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RP	[2]
RN	SEQUENCE FROM N.A.
RA	Duesterhoeft A., Floeth M., Fritz C., Heuss-Neitzel D.,
RA	Hilbert H., Moestl D.;
RL	Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

```
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: SOME, TO HNMI/CTRI.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; U90440; AAB50012.1; -.
DR EMBL; Z71671; CAAG6337.1; -.
DR SGD; S0005339; BIOS.
DR InterPro; IPR002293; AA_rel_permease_1.
KW Transmembrane.
FT TRANSMEM 50 70 POTENTIAL.
FT TRANSMEM 78 98 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 205 225 POTENTIAL.
FT TRANSMEM 231 251 POTENTIAL.
FT TRANSMEM 282 302 POTENTIAL.
FT TRANSMEM 322 342 POTENTIAL.
FT TRANSMEM 370 390 POTENTIAL.
FT TRANSMEM 440 460 POTENTIAL.
FT TRANSMEM 462 482 POTENTIAL.
FT TRANSMEM 508 528 POTENTIAL.
FT TRANSMEM 541 561 POTENTIAL.
SQ SEQUENCE 561 AA; 62428 MW; 7FA5A64FA0D04250F CRC64;
```

```

Query Match      20.1%; Score 540.5; DB 1; Length 561;
Best Local Similarity 25.9%; Pred. No. 1e-27;
Matches 136; Conservative 120; Mismatches 202; Indels 67; Gaps 14;

QY 29 ASESLVPLEKKKOPGTITIVSLAFVICHNSWAGISSLOLALLAGGPVTLLYGILISTLVY 88
   :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 33 SSERSEVKFD-KHFNWWSLGLIAFSLCSGWSGIASMAVGIASGGPLLIYGLIIAAFFS 91
   :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 89 ICIAFSLAETLSVYP-TAGGOYHFASIALPKSI----- 120
   :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 92 LMCIGISLGDAAILPNSSSGGSFWVLKMLEQESVTLKTPEYEDPSDDDEEVLENYCOTFN 151
   :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 121 -----NRSISYVCGLVSLLSWAIGSSVTMPAQOIPALIAAASHYTQSODSWHVFLIY 173
   :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 152 VEVSSFEQRKYSSMVVGULNTFGAIFTTASICSSSMSCIGIHKLHPDYELKHWHVFVG 211
   :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 174 EGVALVLLFNLFALKRNPWHETIGFGLTIALFVISP-IAILARSN-----PKAPNSOV 236
   :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 212 ECINAVLTLPNIYSTPL-PVISQGLVTSLLSFAMTFIICIVSRSDNTVPDPWKASN--I 268
   :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 227 WTAWSNYTGW-SDGVCIFILGSTCFMFIGLDAAHLAEBC---TDAARTVPRXAVVSAILI 283
   :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 269 FGSDMQTGNSSGMFAFVGLNVPDIAFVGIDSATHMIDEVGYSKSRFLVPKRVITIIIV 328
   :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 284 GFCTAFPYTTAVLYGTDLDSILLSSAGYIIP-ETMTQSLRSLSFATVLSCGGIVMAFFAL 342
   :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 329 GFWTSPFYCVGLFFCITDQTAVVES--ILPIVEIFYQATGNRLSVFFLOCNICITTFGVSG 386
   :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 343 NAVQETASRLTWSPARD-----NLGVFTSLHER---IHPRWQVPVWSLPAFWGILATCG 393
   :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 387 IASGTWQSRIOLOSFGKSADPYFKEGSLGNLSKKLAVLTPGFKSPSLYAHFLSQICVITIIG 446
   :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 394 CIFLGSSSTAFNALNVSAVLQQLSFLPIALLLYOKRPKPLPSTRAFV-----LPRGI 447
   :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 447 CIFMGSSTAFNAIITACITILLMSYAVSPFIPLFVIKEKEFIHRIESDVNCVRNRRRM 506
   :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 448 GFLVNVAIVFTSVTVTFSPFLTVPTAASTMNMYTSAIIGVALAL 492
   :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 507 SMIPHICILMTLFCVLFLSPFTPLPVTAGNMNMYTSVVYAVVFCI 551
   :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

```


CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: AF053231; AAC08355.1; -;
DR InterPro: IPR002293; AA_rel_permease_1.
DR InterPro: IPR002027; Amino_acid_permease.
DR Pfam: PF00324; aa_permeases; 1.
DR PROSITE: PS00218; AMINO_ACID_PERMEASE_1; 1.
KW Transport; Amino-acid transport; Transmembrane.
FT TRANSMEM 66 86 POTENTIAL.
FT TRANSMEM 90 110 POTENTIAL.
FT TRANSMEM 139 159 POTENTIAL.
FT TRANSMEM 188 208 POTENTIAL.
FT TRANSMEM 214 234 POTENTIAL.
FT TRANSMEM 255 275 POTENTIAL.
FT TRANSMEM 301 321 POTENTIAL.
FT TRANSMEM 347 367 POTENTIAL.
FT TRANSMEM 399 419 POTENTIAL.
FT TRANSMEM 424 444 POTENTIAL.
FT TRANSMEM 464 484 POTENTIAL.
FT TRANSMEM 496 516 POTENTIAL.
FT SEQUENCE 551 AA; 59869 MW; CB12C646B82F9ADF CRC64;

Query Match 15.2%; Score 409.5; DB 1; Length 551;
Best Local Similarity 26.3%; Pred. No. 2.3e-19;
Matches 140; Conservative 95; Mismatches 236; Indels 61; Gaps 20;

QY 5 PSYGEKGTQRTKNTETAAAGGASELN-VPLEKKQFGTTITIVSLAFVICSNWAGISG 63
DB 25 PSS-GEAASATPSTSTE--SGDKALEALGYTPVFKREFSRSSFSFAVST----SGVYG 77
QY 64 SIQ---LALLAGGPVTLVGLITLVYICIAFSLAELTSVPTAGGVHFAFASILAPKS 119
DB 78 TLMSTWYGLQAGAAAIMWYIGAGGAGWALAYSAEIASAPSSGAMTFTLKFLAPEE 137
QY 120 INRSIVYCVGLVLLSWIAIGSVTMIQAQIPALIAAYSH-TYSQDSWHVFIIYEGVAL 178
DB 138 QVPFLCIAGYLNLVGTVAGGASTEYAAASQMLLAAYVITSNFSVPTPTHVGVMLGTT 197
QY 179 VVLLFNLFALKRNPWHETGFGTITIALFVTSFIALARS-----NPKAPNSQVMTAWSN 232
DB 198 IHAMINTLP---TAWLNRLTSG-----YVVFHISVLLGACVTLVQKRHDHDLKYAFTN 249
QY 233 Y---TGWS-DGVCFILGLSTSCFMFICLDAAMHAECDTAARTVPAKVVSATIIIGFCTA 288
DB 250 FQSSGWSPPGFAFLFGCLTPAWMTGCDTARAEAKNPQMVVPAIANATFTTIVIG 309
QY 289 FPTYIAVLXGITDLSLSAGVIP-----FETMTSLRSLSFATVLSGCGIVMAFFALN 343
DB 310 FFNLVLVCMGPKDLINSQSPQAQLPFNGMRA--PAIFTL--CGFGVNNLVAIP 365
QY 344 AVQETASRLTWSFARNGLVFSTHLERHPRWQVP---VMSLFATWGL-ATCGCIFLGS 399
DB 366 GMO-AGSRTIFALSRLNLLPFSTHWRISKRSTPLIAVW---TYAVLEIILNLLGLAS 420
QY 400 STAFNALVNSAVVLOOLSFILIPALILLYOKRDKPLPSTRAFLVPRIG---FLVNLVAV 456
DB 421 STAIGAVFNCTVALNVYVIPILCKRMVYGRMQG-----PWHMKYSVMVNAFAV 471
QY 457 VFTSVTVTFESFPLTPTAASNTNYSATIGVALGVLNWNVHARKHYOGP 508
DB 472 AWTTFNAVIFFTFTRVPVTPENMNYAIVVFVFLILALVFWYTHGRYYTGP 523

RESULT 7
UGA4_YEAST
ID UGA4_YEAST STANDARD; PRT; 571 AA.

AC P28837;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE GABA-specific permease (GABA-specific transport protein).
GN UGA4 OR YDL210W OR D1037.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SICMA 1278B;
RX MEDLINE=93204891; PubMed=8455553;
RA Andre B., Hein C., Grenson M., Jauniaux J.-C.;
RT "Cloning and expression of the UGA4 gene coding for the inducible
RT GABA-specific transport protein of Saccharomyces cerevisiae.";
RL Mol. Gen. Genet. 237:17-25(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / EY1679;
RA Bahr A., Moeller-Rieker S., Hankeln T., Kraemer C., Schmidt E.R.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR HIGH-AFFINITY, HIGH-SPECIFICITY GABA
CC TRANSPORT.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- INDUCTION: IN THE PRESENCE OF GABA.
CC -!- MISCELLANEOUS: IT REQUIRES NER (NITROGEN PERMEASE REACTIVATOR
CC PROTEIN) FOR ITS FULL ACTIVITY.
CC -!- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X66472; CAA47101.1; -;
CC EMBL: X99000; CAA67481.1; -;
CC EMBL: Y74258; CAA98788.1; -;
CC PIR: S30253; S30253.
CC SGD: S0002369; UGA4.
CC InterPro: IPR002293; AA_rel_permease_1.
CC InterPro: IPR002027; Amino_acid_permease.
CC Pfam: PF00324; aa_permeases; 1.
CC PROSITE: PS00218; AMINO_ACID_PERMEASE_1; 1.
KW Transport; Transmembrane.
KW SEQUENCE 571 AA; 61872 MW; 051B5828654EC784 CRC64;

Query Match 14.2%; Score 383.5; DB 1; Length 571;
Best Local Similarity 25.8%; Pred. No. 1.1e-17;
Matches 142; Conservative 107; Mismatches 236; Indels 65; Gaps 22;

QY 6 SGYGEKG--GTRQTKNTETAAAGSSELNVPLEKKQFGTTITIVSLAFVICSNWAGISG 63
DB 41 TGAGEVNVYDAKSVNDNQLLAIEIGYKQEL-----KRFSTLQVFGIAFSIMGLPSIA- 94
QY 64 SLQALLAGGPVTLVGLITLVYICIAFSLAELTSVPTAGGVHFAFASILAPKSINRS 123
DB 95 SVMGGGLGGPATLVWGFVAFAFFILLVGLITWAEIHASSIPTAGGLYWTYYTAPEGYKEI 154
QY 124 ISYVGLVSLLSWIAIGSVTMIQAQIPALIAAYSHYSQD-----SWHVFIIYEGVA 177
DB 155 ISFIIGSNSLALAAAGVCSIDYGLAEATAAV-----TLTKDGNFVTSGLKLYGFAGAV 209
QY 178 LVVLLFNLFALKRNPWHETGFGTITIALFVTSFIALARSNPKA-----PNSQVMTAWSN 233
DB 210 VVMCTCTCVASGAIARLQTLTSIFANLFIIVLLFIALLPIGTRKRMGGFNDGDFYGYENL 269
QY 234 TGWSDGVCFIL-GLSTSCFMFICLDAAMHAECDTAARTVPAKVVSATIIIGFCTAPPT 292

Db 270 SDNNNQWQCLAGFPAVMTGSDSCVHSEAKDAKSKYPIGIISSIAVCWILGLWII 329
QY 293 IAVLYGIT-DLDSILSA-CYIPETMTQSLR---SLSFATVLSGGVTVMAFFALNAVQE 347
Db 330 ICLMACINPDIDSVLSKYGAFALAIYDLSGKKWAFAMSL-----TAFCOFLMGASIT 384
QY 348 TA-SRLTWSFARDNGLVSTHLRIHPRWQVPVMSLFAFMGILATC-GCIFLG-----SS 400
Db 385 TAVSRQVMAFSDRGLPLSKYIKRVDKSYVPFFA-----ILAACVGSLLILGLCLLIDD 438
QY 401 TAFNALVNAVVLQSLFLPIALPIALLLYQKRPKLPSTRAFLVPRGIGFLVNLAVVFTS 460
Db 439 AATDALFLAVAGNLANWSTPTFRLTSGRD-LFRPG--PFYLGKWSPIVAWTGVAFQL 495
QY 461 VTVTFSEF-----LVPTAASMTNYTSAIIGVAL-ALGVNLVWVHARKHYOGPHLELDG 514
Db 496 FIILWVFFSQHGII-----KSTMYN-ACVIGPQINILAGIYKYKHYHGPATNLS 550
QY 515 ----RVVGAE 520
Db 551 DDVTEAVGAD 560

RESULT 8

YA08_SCHPO STANDARD; PRT; 550 AA.
AC Q10087;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative amino-acid permease C11D3.08c.
GN SPAC11D3.08c.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 268166; CAA92309.1; -
DR InterPro; IPR002293; AA_rel_permease-1.
DR InterPro; IPR002027; Amino_acid_permease.
DR Pfam; PF00324; aa_permeases; 1.
DR PROSITE; PS00218; AMINO_ACID_PERMEASE_1; 1.
KW Hypothetical protein; Transport; Amino-acid transport; Transmembrane.
FT TRANSMEM 53 73 POTENTIAL.
FT TRANSMEM 78 98 POTENTIAL.
FT TRANSMEM 127 147 POTENTIAL.
FT TRANSMEM 177 197 POTENTIAL.
FT TRANSMEM 201 221 POTENTIAL.
FT TRANSMEM 246 266 POTENTIAL.
FT TRANSMEM 283 303 POTENTIAL.
FT TRANSMEM 336 356 POTENTIAL.
FT TRANSMEM 386 406 POTENTIAL.
FT TRANSMEM 411 431 POTENTIAL.
FT TRANSMEM 453 473 POTENTIAL.
FT TRANSMEM 481 501 POTENTIAL.
SQ SEQUENCE 550 AA; 59918 MW; 3DD1EF6003896B4C CRC64;

Query Match 13.1%; Score 351.5; DB 1; Length 550;
Best Local Similarity 23.4%; Pred. No. 1.2e-15;
Matches 129; Conservative 107; Mismatches 238; Indels 77; Gaps 16;
QY 10 EKGTROTTNTETAAGGASESLN-----VPLEKKQFGTITVSLAFVICNSW 58
Db 5 DKESLKQKQWENEVELGESREANQDDELLMSLGYPKPEFTREFSYVSIFGQSGMGLC 64
QY 59 AGTSGSLQALLAGGPVTLTYGILISTLVYICIAFSLAELTSVYPAGGOYHFAISLAPK 118
Db 65 PAMAGSLIFSNCGGG-GMWWSWIIIGICLIPVSLGELASMPSTGGGLYFIWFTLASP 123
QY 119 SINRSIVVCGVLSLLSWIAIGSSVTMIPAOQIPALIAAISHYTSODSWHVFLLIEGVAL 178
Db 124 SSRFLCWGCVSVGLVATYVASTVYSSSMVQALVIGSPSYSTKYEQYGYAALLF 183
QY 179 VVLLFNLFALKRNPWVHBIIGFLITIALFVISFIALARSNKAPNSQ--VWTAMSNYTW 236
Db 184 VISAMTAIPSRVIAKVINIITFQFLVSLIILIAAAGSDSTRNSGSIFFGDTNYSGW 243
QY 237 SD-GVCFILGLSTSCFMFGLDAAHMLAEECTDAARTVPKAVS----ATIIIGCTAF 291
Db 244 SNMGWAFILSFTTPVWVVSFGFESSAAVAEESTNAARAPFAMISSLGATILGWCIIV-- 301
QY 292 TIAVLYGITDLSIL-SSAGYIPFETMTQSLRS-----LSFATVLSGGVTVMAFFA 341
Db 302 TVVATMG-HDFNAILGSSLGQPVAGVLYNNVGNKAGLIFSLVIALCLNCISLLIA--- 357
QY 342 LNAVQETASRLTWSFARDNGLVSTHLRIHPRWQVPVMSLFAFMGILATCGCIFLGSS 401
Db 358 -----ASREVFACRDCGGIPGSRYL-RLLTQKQVPLNAILLVLLYSLVGLLILVNV 409
QY 402 AFNALVNSAVVQLQSLFLPI-ALLLYQKRPKLPSTRAFLVPRGIGFL-----VNV 454
Db 410 AISSVFNLAIIALYIAYSGPLMCRFVYKFKP-----GVFVGVKSKKPAAL 456
QY 455 AVVFTSVTVTFSPFLTPVPTAASMTNYTSAIIGVALALGVNLVWVHARKHYOGPHLELD 514
Db 457 SLVWVWFEMILMLLPQYQKPNQDENWALVILGVVWFCVV-----YYLIP----- 502
QY 515 RVVGAEFQVGP 525
Db 503 KIGGKTEFTGP 513
RESULT 9
GPTI_CANAL STANDARD; PRT; 553 AA.
AC 074248;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Putative polyamine transporter.
GN GPTI.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RA McNemar M.D., Gorman J.A., Buckley H.R.;
RT "Isolation and sequence of the GPT1 gene of Candida albicans, encoding
RT a putative polyamine transporter";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; AF080132; AAC31569.1; -
DR InterPro: IPR002293; AA_rel_permease_1.
DR InterPro: IPR002027; Amino_acid_permease.
DR Pfam; PF00324; aa_permeases; 1.
DR PROSITE; PS00218; AMINO_ACID_PERMEASE_1; FALSE_NEG.
KW Transport; Transmembrane.
FT TRANSMEM 43 63 POTENTIAL.
FT TRANSMEM 68 88 POTENTIAL.
FT TRANSMEM 133 153 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
FT TRANSMEM 190 210 POTENTIAL.
FT TRANSMEM 240 260 POTENTIAL.
FT TRANSMEM 280 300 POTENTIAL.
FT TRANSMEM 333 353 POTENTIAL.
FT TRANSMEM 397 417 POTENTIAL.
FT TRANSMEM 453 473 POTENTIAL.
FT TRANSMEM 485 505 POTENTIAL.
SQ SEQUENCE 553 AA; 60633 MW; 197DACB76C5FBC9E CRC64;

Query Match 12.4%; Score 334; DB 1; Length 553;
Best Local Similarity 23.5%; Pred. No. 1.6e-14;
Matches 120; Conservative 91; Mismatches 233; Indels 56; Gaps 18;
Qy 39 KQKQFTITIVLAFCVNSWAGISGLQALAGGPVTLVYIGILSTLVYICIAFSLAEL 98
Db 37 RREFSLWSIFAVSFVGLLPSIAACFDYQQLVGVMSPLPW-LIAMFITSVAYSMABI 94
Qy 99 TSVYPTAGQYHFAISILAPKSNRSISVCGVLVSLSWIAIGSSVTMTIPAQIPALIAAY 158
Db 95 ASAFPSCAGTPYAVSOLAPK-----KYASFLTWTCTWNWSCQTAAPS-----VSY 141
Qy 159 S-----HNYSDSW-----HVLIVGEVALVLLFNLFLKRNPMV---HEIGFGLT 202
Db 142 SCACMLALHSPTDFSVASNAQIFGLTGIQVLCAFMACFTPK---WVAFSSAGTTCTN 198
Qy 203 IALFVISTAILA-----RSNPKA-----PNSQVWTAWSNYTCWSDGVCFTILGLSTSCFMF 253
Db 199 IVFLVVFVIMILGKNKQDIKEIGISKFNSNSTAW-GLDNOAEWPTGLSFLISFMGVIAW 257
Qy 254 IGLDAMHIAECTDAARTVPAVSAIIGCTAPPTIYAVLYGTDLDSILS---SAG 310
Db 258 SGYDSPFHAEKCSNAVAAPRAIVLTSTVGLGICGFEMFIAIAYTLVLDNQISADPESLG 317
Qy 311 --YIPETMTQSLRSIPATVLSGCGIVMAFFALNAVOETASRLTWSFARDNGLVFSTHL 368
Db 318 QPFTVLTQIMDKNLVIGATALT---IISFFMAQNCILLASREVYAYARDGLFPLSGIW 374
Qy 369 ERIHPRWQVPVWSLFATWGILATCGCIFGSGTAFNALNSAVVLOOLSFLLPIALLLYQ 428
Db 375 KKVSPKTOPPINAVIMNEFVEELLILLIFRGDVSIGSIFSIGALAGFISFTMTPLKITY 434
Qy 429 KDDPFLPSTRAFLPRGICFL---VNVLAIVFTSTVTFEFPPLVPTAAST---MNYT 482
Db 435 AR-KTFQPG-----PWNIGKWESEPGWVSVAFVGLMVLPILCEP-TVKGADLPTTEMNWT 486
Qy 483 SAIGVALALGVLNWWVHARKHQGPHELE 512
Db 487 CLVYFGLILLTTFVWVYARRWVGPRTWI 516

RESULT 10
YK4_YEAST
ID YK4_YEAST STANDARD; PRT; 618 AA.
AC P36029;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical amino-acid permease in STE3-GIN10 intergenic region.

GN YK174C OR YK1639.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=94378719; PubMed=8091858;
RT Vandenbol M., Bolle P.-A., Dion C., Portetelle D., Hilger F.;
RT "Sequencing and analysis of a 20.5 kb DNA segment located on the left
RT arm of yeast chromosome XI.";
RL Yeast 10:S25-S33(1994).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; Z26878; CAAB1512.1; -
DR EMBL; Z28174; CAAB2016.1; -
DR PIR; S38004; S38004.
DR PIR; S38406; S38406.
DR PIR; S44586; S44586.
DR SGD; S0001657; YK174C.
DR InterPro: IPR002293; AA_rel_permease_1.
DR InterPro: IPR002027; Amino_acid_permease.
DR Pfam; PF00324; aa_permeases; 1.
DR PROSITE; PS00218; AMINO_ACID_PERMEASE_1; 1.
KW Hypothetical protein; transport; Amino-acid transport; Transmembrane.
FT TRANSMEM 61 84 POTENTIAL.
FT TRANSMEM 91 110 POTENTIAL.
FT TRANSMEM 132 148 POTENTIAL.
FT TRANSMEM 155 171 POTENTIAL.
FT TRANSMEM 180 200 POTENTIAL.
FT TRANSMEM 212 231 POTENTIAL.
FT TRANSMEM 298 317 POTENTIAL.
FT TRANSMEM 343 367 POTENTIAL.
FT TRANSMEM 403 419 POTENTIAL.
FT TRANSMEM 426 449 POTENTIAL.
FT TRANSMEM 465 486 POTENTIAL.
FT TRANSMEM 499 516 POTENTIAL.
SQ SEQUENCE 618 AA; 69242 MW; 035CB927C624CE74 CRC64;

Query Match 11.9%; Score 321; DB 1; Length 618;
Best Local Similarity 24.6%; Pred. No. 1.2e-13;
Matches 116; Conservative 99; Mismatches 208; Indels 48; Gaps 15;
Qy 37 LEKQKQFTITIVLAFCVNSWAGISGLQALAGGPVTLVYIGILSTLVYICIAF--- 93
Db 51 LDKSLLSRGSTVGLGLMSPLVGMCTSMACGLINGSPGLTIMGFLISG---VCWFSSL 107
Qy 94 SLAELTSVYPTAGQYHFAISILAPKSNRSISVCGVLVSLSWIAIGSSVTMTIPAQIP 152
Db 108 SLGEIVSKFPM---ELHVGSAMLAPEKUKLVCSWTGMLNLGNWTNSTSIITFAGALTI 164
Qy 153 ALIAAHSHTYSQDSWHVF-----LIYEGVALVLLFNLFLKRNKRP-----WVHEIGFGL 201
Db 165 SLILMTNSNLISEAHLIFVTYVFLVTVVGLVNLKFAFRTETINKVCVYI----- 217
Qy 202 TIALFVISTAILA-----RSNPKAPNSQVWTAWSNYTCWSDG-VCFILGLSTSCFMFICL 256
Db 218 ---IVAIIFIDILLVLFHKGKFRSLKVALFHFDNNLSGYKSAFLSFIIGFQOSNFTLQGF 274
Qy 257 DAAMHIAECTDAARTVPAVSAIIGCTAPPTIYAVLYGTDLDSILSSAGVTP--- 313
Db 275 SMLPALADEVKVPEKIDIPRGMNAVLLSAFSGVIFLIPMLILPDNDLLFTNHKVLPIVN 334

QY 314 -FETWTSRLSLSPATVLSGGIFLWAFNALNAVOETASRLTWSFARDNGLVFTSLERIH 372
 Db 335 IFTSTDSVLSFFLVLLILGNLL--FSGIGSI--TTSRAVYSRDRQAIPYDKWTVE 391
 QY 373 PRWQ--VPVWLSFATWGLATCGCIFLGSSAFNALNSAVVLOQLSLIPIALLLYOKR 430
 Db 392 PDSQKVPKNSVWLSMIISYFLGLLALISTAFAENAFIGAVALCLCSATFIPLVLFVTR 451
 QY 431 DPKFLPSTRAFLVPRGIFLVNVLAVFTSVTTVFSSPLPLVPTAASTMYN 481
 Db 452 --RAIRAPVKIRYK-FGWFINIVISVWLLSLMSVSVCLPTQVPTKTMNY 499

RESULT 11
 YFYA_SCHPO
 ID YFYA_SCHPO STANDARD; PRT; 554 AA.
 AC Q59813; P78767;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative amino-acid permease C794.03.
 GN SPCC794.03.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Murphy L., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 198-554 FROM N.A.
 RC STRAIN=PR745;
 RX MEDLINE=98162722; PubMed=9501991;
 RA Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;
 RT "Identification of open reading frames in Schizosaccharomyces pombe cDNAs."
 RL DNA Res. 4:363-369(1997).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AL023595; CAA19131.1; -
 DR EMBL; D89115; BAA13777.1;
 DR InterPro; IPR002293; AA_rel_permease_1.
 DR InterPro; IPR002027; Amino_acid_permease.
 DR Pfam; PF00324; aa_permeases; 1.
 DR PROSITE; PS00218; AMINO_ACID_PERMEASE_1; FALSE_NEG.
 KW Hypothetical protein; Transport; Amino-acid transport; Transmembrane.
 FT TRANSMEM 68 88
 FT TRANSMEM 126 146
 FT TRANSMEM 176 196
 FT TRANSMEM 206 226
 FT TRANSMEM 288 308
 FT TRANSMEM 397 417
 FT TRANSMEM 492 512
 FT TRANSMEM 523 593
 FT CONFLICT 293 293 A -> D (IN REF. 2).
 FT CONFLICT 305 308 VMLF -> RHVEI (IN REF. 2).
 FT CONFLICT 313 313 MISSING (IN REF. 2).
 FT CONFLICT 372 372 A -> G (IN REF. 2).
 FT CONFLICT 412 412 P -> A (IN REF. 2).
 FT CONFLICT 423 423 A -> D (IN REF. 2).
 SQ SEQUENCE 554 AA; 60857 MW; F88F245437F1265C CRC64;

Query Match 11.6%; Score 311.5; DB 1; Length 554;
 Best Local Similarity 25.6%; Pred. NO. 4.3e-13;
 Matches 121; Conservative 88; Mismatches 221; Indels 43; Gaps 17;
 QY 51 AFVICSNWAGISGSLQALLAGGPVTLVGLILSTLVYICIAFLAELTSTVYPAGGOYH 110
 Db 54 SFAACDCMSNIRGSFYIGLLTGCPSSAYITIIAIPQLISAAATMAEVCALPTAGSLYF 113
 QY 111 FASILAPKINSISYVCG---LVSLLSWIAIG--SSVTMIPAAQOIPALIAAYSHYSYSDS 166
 Db 114 WASAAGGKYGRLGFIWAVWVAVTSTFVAVNCQSTTFEIGELPVFNSGFSVSSDOVK 173
 QY 167 WHV--FLIYEGVALVLLFNLFALKRNPWVHEIGFGLTIALFVIFISIAI--LARSNPKAPN 223
 Db 174 FRAVQWAVEAILLVCVLLNFIPPKWFRVFRVSVAIVLLDFVLNMIWLPVAVSTKYGR 233
 QY 224 SQWWTANSY-----TGHSDGVCFILGLSTSCFMFEGIDAAHMLAEECTDAARTVPK-- 275
 Db 234 DEAFMKSTNYDLGKVNNGWS----WCLTFECTARILVGYDAAGHVAEETKNASKTSRGM 289
 QY 276 ---AVVSAIL-IGFCTAPPYTIIVLYGITDLSILSSAGVIPETWTQSLRSLSFATVLS 331
 Db 290 FYSAFSAIILSTGLIIVMFLYCLPPSNVWYELIKNSQPFVSYAYALGRAHVMNVV- 348
 QY 332 CGGIVMAFFALNAVOETASRLTWSFARDNGLVFTSLERIHPRWQVPMVMSLFATWGLAT 391
 Db 349 --GILGMIFDTSLSIVASSRLVFAVARDGVLPFSGWLKRVDSHGQ-PTNAVTFILISAA 405
 QY 392 CGCIFLGSSSTAFNALNSAVVLOQLSF-LIPIALLLYOKRD-PKFLPS-----TRAVLPR 445
 Db 406 LLCNLPSAFAPTSLLSAAAVPTIMAYAAVAFGRFLSLRDPKSEWSLGSKPFQL-- 463
 QY 446 GIGFLVNLAVVFTSVTTVFSSPLTVPPTAASTMYNTSAIIGVALGALGVLMWV 498
 Db 464 -ITFLWN---LFTAV--ILFS-PKAYPTGKNFVNPVPIFGAITIFGLISWL 508
 RESULT 12
 YFYA_SCHPO
 ID YFYA_SCHPO STANDARD; PRT; 591 AA.
 AC Q9UT18;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative amino-acid permease C9.10.
 GN SPAC9.10.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Wedler H., Duesterhoeft A., Lyne M.H., Rajandream M.A., Barrell B.G.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AL121764; CAB57428.1; -
 DR InterPro; IPR002293; AA_rel_permease_1.
 DR InterPro; IPR002027; Amino_acid_permease.
 DR Pfam; PF00324; aa_permeases; 1.

Db 152 SYNLVANFWTLVLVQISIMVVFLLVVOGLHKGEGVGTWMSLQPFISENAHLIPIITGA 211
 QY 247 STSCFMFGLDAMHLAECDTAARTVPKAVVSAILIIGFCTAPPTTYTAVLYGTDLDSIL 306
 Db 212 TIVCFSEFLGDAVTTTSEETDAARVIPKAI-----FLTAVYGGVIFTAASF 258
 QY 307 SSAGVIP-----FETWQSLRSL-----FATVLSGGGIVMAFFALNAVQETASRLTWS 355
 Db 259 FMQLFFPDISRKDPDAALPIALYVVGKLFQSFILCTFTFNTLASGLASHASVSRLLV 318
 QY 356 FARDNGLVFSTHLER----IHPRQVQWWSLFAWTGILATCGCIF-LGSSTA---FNALV 407
 Db 319 MGRDN--VFP---ERVFGYVHPKRWTPALNAVIMV-GIVALSALFFDLVTATALINEGALV 372

 RESULT 14
 ID YEEF_ECOLI STANDARD; PRT; 452 AA.
 AC P33016;
 DT 01-PEB-1994 (Rel. 28, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical transport protein yeeF.
 GN YEEF OR B2014 OR Z3176 OR ECS2816.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / BHB2600;
 RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
 RA Church G.M.;
 RL Submitted (OCT-1993) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=97251358; PubMed=9097040;
 RA Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,
 RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
 RA Saito N., Sempel G., Seki Y., Sivasubram S., Tagami H.,
 RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
 RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 40.1-50.0 min region on the linkage map.";
 RL DNA Res. 3:379-392(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11208551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouisis K.,
 RA Apodaca J., Anantharam T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=12074935; PubMed=11208551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouisis K.,
 RA Apodaca J., Anantharam T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 RN [5]


```
QY 285 PCTAPPYTIATVLYGTTDLSDLSLSSAG-----YIPFETMTQSLRSLSFATVLSGCGIVM 337
Db 247 MIFIFATYFLQLK-FPDISRFKDPDASQPEIMLYVAGKAFQVGLIFSTIIVLASSG----- 301
QY 338 AFFALNAVQETASRLTWSFARDNGLVFSTHLERIHPRWQVPVWSLFAFWGI----- 388
Db 302 -----MAAHAGVAKLMVWGRD-GVFPKSFPGYVHPKMTTPAMNIIYLGATALLAINPDL 355
QY 389 -LAPCGCIGFLSGSTAFNALNSAVLQQLSELPIATLLYQKRDPK-----PLPSTR 439
Db 356 VMAT-ALINFALGAVAF-TFVNLSVISO-----FWIREKRNKTLKDHQFLPLPCMG 404
QY 440 AFVLPKRGIGEL-VN-----VLAVVETSVTVTF-----SEPLTPV 473
Db 405 ALT-----VGALWNLESSVNLGAIWAIIGLIYLVACVTKSFRNPV 446

RESULT 15
YJ99_MYCTU
ID YJ99_MYCTU STANDARD; PRT; 440 AA.
AC Q10858;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 45.3 kDa transport protein Rv1999c.
GN Rv1999c OR MT2055 OR MTCY39.19.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID:1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.F., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLE AMINO-ACID OR METABOLITE TRANSPORT PROTEIN.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR InterPro; IPR002422; AA_rel_permease_2.
DR InterPro; IPR002027; Amino_acid_permease.
DR Pfam; PF00324; aa_permeases; 1.
DR PROSITE; PS00218; AMINO_ACID_PERMEASE_1; FALSE_NEG.
KW Hypothetical protein; Transport; Transmembrane; Complete proteome.
FT TRANSMEM 24 44 POTENTIAL.
FT TRANSMEM 47 67 POTENTIAL.
FT TRANSMEM 93 113 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 155 175 POTENTIAL.
FT TRANSMEM 183 203 POTENTIAL.
FT TRANSMEM 229 249 POTENTIAL.
FT TRANSMEM 276 296 POTENTIAL.
FT TRANSMEM 323 343 POTENTIAL.
FT TRANSMEM 346 366 POTENTIAL.
FT TRANSMEM 379 399 POTENTIAL.
FT TRANSMEM 400 420 POTENTIAL.
SQ SEQUENCE 440 AA; 45267 MW; 9592E0A842D18CD CRC64;

Query Match 7.4%; Score 199.5; DB 1; Length 440;
Best Local Similarity 23.9%; Pred. No. 5e-06;
Matches 117; Conservative 75; Mismatches 179; Indels 119; Gaps 25;

QY 34 NVPLE-KKQFGTI--TIVSLAFVIONSWAGISGSLQALLAGGPVTLTYGLIISTLVYIC 90
Db 9 DIPDELRRRLGLLDVAVVIGIGSMI---GAGIFAALAPAPAYAGS-GLILGLAVAAVAYC 64
QY 91 IAFSLAELTSVYTAGQYHFASILAPKPSINRSISVCGVLSSLWS-IAIGSSVTMIPAQ 149
Db 65 NATSSARLAARYPASGGTYVYGERM---RLGDFWGLAG-----WGFVVGKT----- 107
QY 150 OIPALIAAYSHYTSQDSNHWFLIYEGVALVLL--FNLPALKRNPVW-HEIGFGLTIALP 206
Db 108 ---ASCAAMALAVGFTVMPAQAHAVAVYVAALTAIVNAGIOKSAWLTRSI---VAVVLV 161
QY 207 VISFIAIATLARSNPKAP-----NSQVWTAMSNYTGSDGVCFFILGLSTSCFMFIGLDA 258
Db 162 VLTAVVVAAYGSGAADPARLDIGVDAHVV-----GMLQAAGL--LFFAFAGVAV 208
QY 259 AMHLAECTDAARTVPKAVVSAIIIGFCFAPYTTAVLYGITDLSILSSAGYIPFETMT 318
Db 209 IATLGBEVRDPARTIPRAIPLALGI---TLAVTAVAVAVIAVLGPQLARAAP---LS 262
QY 319 QSLRSLS---FATVLSGCGIVMAFFALNAVQETASRLTWSFARDNGLVFSTHLER----I 371
Db 263 EAMRVAGVNWLIIPVWQIGAAVAALGSLALILGVSRTTLAMARDR-----HLPRWLA 316
QY 372 HPRWQVP-----VMSLEATWGLATCCGIFLGSSTAFNALVNSAVVLQLSFLIP 421
Db 317 HPRKVPFFRAELVVGAVAAALATAADIRGAIG-----FSSFGVLYVYAIANAS-----A 365
QY 422 IALLLYOKRDKPELKFSTRAPVLPKRGIFGLVNLVAVVFTSVTTVFFSFPLTVPVTA 481
Db 366 LTGLDEGRPRRLIP-----LVGLIGCV-----VLAFALPLSSVAAG----- 402
QY 482 TSAIIGVALA 491
Db 403 -AAVLGVGVA 411
```

Search completed: April 27, 2002, 07:59:41

Job time: 747 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:48:21 ; Search time 68.15 seconds
(without alignments)
740.233 Million cell updates/sec

Title: US-09-882-694A-8

Perfect score: 2692

Sequence: 1 MDSRPSGYGEGKGTROTAKN.....QGPHELDGRVGAEFQVGP 525

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	681.5	25.3	563	1 S11175	choline transport
2	540.5	20.1	561	2 S63388	probable membrane
3	454	16.9	567	2 T50051	amino-acid permeas
4	451	16.8	517	2 D84421	probable amino aci
5	445	16.5	542	2 T39474	amino acid permeas
6	429	15.9	557	2 T41588	probable amino-aci
7	420	15.6	544	2 S62527	probable amino-aci
8	383.5	14.2	571	1 S30253	GABA transport pro
9	351.5	13.1	550	2 T37519	probable transport
10	321	11.9	618	1 S38004	probable amino aci
11	311.5	11.6	554	2 T41612	amino acid permeas
12	274	10.2	591	2 T39195	probable amino aci
13	243.5	9.0	496	2 B83503	probable amino aci
14	227.5	8.5	465	2 E69825	amino acid transpo
15	226.5	8.4	479	2 C64878	probable amino aci
16	226.5	8.4	479	2 A90863	probable amino aci
17	225.5	8.4	461	2 D69814	metabolite transpo
18	225.5	8.4	461	2 H85755	probable amino aci
19	225	8.4	463	2 AH0189	Amino acid permeas
20	222.5	8.3	454	2 AG0763	probable amino aci
21	221	8.2	776	2 A84178	cationic amino aci
22	218	8.1	440	2 G83195	probable amino aci
23	216	8.0	482	2 C90067	hypothetical prote
24	215.5	8.0	454	2 E64966	probable amino aci
25	215.5	8.0	454	2 H90980	probable amino aci
26	215.5	8.0	454	2 E85826	probable amino aci
27	210.5	7.8	456	2 B83391	probable amino aci
28	202.5	7.5	494	2 B89827	hypothetical prote
29	201	7.5	463	2 AE1155	amino acid transpo

30	199.5	7.4	440	2 E0758	hypothetical prote
31	199	7.4	482	2 G83142	probable transport
32	199	7.4	556	2 G70503	probable cycA prot
33	197.5	7.3	596	2 S46001	probable amino aci
34	196	7.3	539	2 D69748	amino acid transpo
35	195	7.2	447	2 AI0701	probable amino aci
36	194	7.2	647	2 E90259	amino acid transpo
37	193	7.2	465	2 AG0415	aromatic amino aci
38	191.5	7.1	736	2 C69451	cationic amino aci
39	189.5	7.0	463	2 AH1758	amino acid transpo
40	189.5	7.0	496	2 G83136	probable amino aci
41	189.5	7.0	504	2 AG2839	hypothetical prote
42	189.5	7.0	518	2 A97617	probable transport
43	189	7.0	455	2 F83460	probable amino aci
44	189	7.0	465	2 A83046	probable amino aci
45	188.5	7.0	452	2 G95306	putrescine/ornithi

ALIGNMENTS

RESULT 1

S11175

choline transport protein - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein G3213; protein YGL077c

C:Species: Saccharomyces cerevisiae

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000

C:Accession: S11175; S64084

R:Nikawa, J.I.; Hosaka, K.; Tsukagoshi, Y.; Yamashita, S.

J. Biol. Chem. 265, 15996-16003, 1990

A:Title: Primary structure of the yeast choline transport gene and regulation of its

A:Reference number: S11175; MUID:90368823

A:Accession: S11175

A:Molecule type: DNA

A:Residues: 1-563 <NIK>

A:Cross-references: EMBL:J05603; NID:g171329; PIDN:AAA34537.1; PID:g171330

R:Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S64071

A:Accession: S64084

A:Molecule type: DNA

A:Residues: 1-563 <RIE>

A:Cross-references: EMBL:z72599; NID:g1322592; PIDN:CAA96782.1; PID:g1322593; GSPDB:C

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:HNMI; CTR1; MIPS:YGL077c

A:Cross-references: MIPS:YGL077c; SGD:S0003045

A:Map position: 7L

C:Superfamily: choline transport protein

C:Keywords: transmembrane protein

F:91-107/Domain: transmembrane #status predicted <TM1>

F:185-201/Domain: transmembrane #status predicted <TM2>

F:214-230/Domain: transmembrane #status predicted <TM3>

F:257-273/Domain: transmembrane #status predicted <TM4>

F:298-314/Domain: transmembrane #status predicted <TM5>

F:345-361/Domain: transmembrane #status predicted <TM6>

F:401-417/Domain: transmembrane #status predicted <TM7>

F:428-444/Domain: transmembrane #status predicted <TM8>

F:466-482/Domain: transmembrane #status predicted <TM9>

Query Match

Best Local Similarity 25.3%; Score 681.5; DB 1; Length 563;

Matches 156; Conservative 95; Mismatches 210; Indels 9; Gaps 6;

QY 39 KKQFGTTIVSLAFVNCNSWAGISGSLQALLAGGPVTLVGLISLTVVICIAFSLAEL 98

Db 55 RKFSFLWSILGVGFGITNSWFGISTSWAGISSGPGMWIVGIIVIALISICITSGEL 114

QY 99 TSYPTAGQYHFASIIAPKSIINRSISYVGLVSLLSWIAIGSSVTMPAQOIPALIAAY 158

Db 115 SSAYPHAGGFWMSLKLPKRYKRFAYAGSVFTSASTTILSVATEVVGMYALT 174


```

317 MQIIIVAVTLTIDSVMMNTSGSMWTAYLVQAPPKAALGILSLTIIISAIIMGOSALIA--- 373
348 TASRLTWSFARDNGLVFSTHLERIHPRWQYVPVNSLFTATGILATCGCI-----FLG 398
374 -SSRIASYARDGTLPSGWIQGVNPTQPVNAVI-----CNCISILILFTTFAG 424
399 SSTAFNALVNSAVVLOQLSPILDIPTALLLYOKRDKPLPSTRAVLPRGIGFL---VNVLA 455
425 TVT-LDAVFSVGAAVAFIATVPIATRVFTKDAEFRG-----PWNLGKFSRPIGLLA 477
456 VVFTSVTTVFSPRLTPTTAASMTNNTSATIGVALALGVNLVNVVHARKKHVQGP 508
478 VSFVALMIPILCPFSVKNPTAQEMNTCLVYGGPMLFTLVWYISARKWFKGP 530

RESULT 4
D84421
probable amino acid permease [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2000
R:Accession: D84421
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis th
A:Reference number: A84420; MOID:20083487
A:Accession: D84421
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-517 <SPO>
A:Cross-references: GB:AE002093; NID:g4262224; PIDN:RAD14517.1; GSPDB:GN00
C:Genetics:
A:Gene: At2g01170
A:Map position: 2
C:Superfamily: choline transport protein

```

[illegible]

```
QY      493  GVLNWVVHARKHYQGP 508
          :  | :  | :  | :  |
Db      496  TLSYWLFSARHWFTGP 511
```

RESULT 5
T39474
amino acid permease - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T39474
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Xiang, Z.; Aves, S.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z11857
A:Accession: T39474
A:Status: preliminary; translated from GE/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-542 <LYN>
A:Cross-references: EMBL:AL023290; PIDN:CAA18895.1; GSPDB:GN00067; SPDB:SPBC15C4
A:Experimental source: strain 972h-; cosmid c15C4
C:Genetics:
A:Gene: SPDB:SPBC15C4.04c
A:Map position: 2
C:Superfamily: choline transport protein

Query Match	16.5%	Score 445;	DB 2;	Length 542;
Best Local Similarity	29.3%	Pred. No. 1.2e-24;		
Matches	146;	Conservative 83;	Mismatches 229;	Indels 40; Gaps 16;
Qy	36	PLEKKQFGCTIVIVSLAFVICSNWAGISGSIQALLAGGQVPTLLYGILLISTLIVYICIAFSL	95	
Db	55	VPFKREFSTWATFSFAGISGLFATVTVTYSYPLISGAGPAAVNCWLGACAGCMCAISLV	114	
Qy	96	AELSVYPTAGGQYHFASILAPKPSINRSYVCGVLVSLLSIAIGSSVTMIPAAQIQIPALI	155	
Db	115	AELVSAYPTSGGLYFTCKDLVPARSPVAVVWGLNLGGQAAGVSSSTDWSCAQ---	171	
Qy	156	AAYSHT-----YSQDSWHVFLYEGVALVVLFFN-LFALKRNPWVHEI--GFGLT--IALFV	207	
Db	172	AAVSISTDLKYPTNQHIV----GVMAAAIVVFHGLVNSLSLWLDRIREFVATFHLIVLV	227	
Qy	208	ISFTAILARSPKAPNSQVWTAWSNYTCWSD-GVCFILGLSTCFMFILGLDAAMHLAECC	266	
Db	228	VMCICLLAKCPKFNTRYFTDQASSGWHPIGFSFUGFLSVAMCWTDYDATAHIAEEI	287	
Qy	267	TDARTVPKAVVSATIIIGCFATFPVTIAVLXGI-TDLSILSSAGYIP---FETMTQSL	321	
Db	288	ENAAVRAAPNALALSTIYVLCWFENIVLATFTMGTDLDLSINSELGOPVAQIIFYNVLGKK	347	
Qy	322	KSLSFATVLSGGIVMAFFALNAVQETASRLTWSFARDNGLVFSTHLERIHPRMOVVPWS	381	
Db	348	GSMAP-TILSF--IIINFTGITAMQANA-RTIWAFSRDQALPFSRYWYKINKTITPTV--	401	
Qy	382	LPATWGIILATC---GCIFLGSSTAFNALVNSAVVLOQLSFLIPIATLLYQKRDPKFLPST	438	
Db	402	-IAVWLVNVFCIALNLIGLSIEAIEALFSVCAIALDWSYVIPACKLIFGKRLNYKPG-	459	
Qy	439	RAFVLPGRGIGF---LVNVYLVVFTSVTVTFVSFPLTVPVTAASTNNYTSIIAGTVALACVL	495	
Db	460	-----PWNLGWASHFVNAYACWCTAFVSVIFELMPTVRPVTQNNMYAVVVLGAVLLSLV	514	
Qy	496	NWVHARKHYOGPHLELD	513	
Db	515	YVWWSGARKSYICPRINVD	532	

RESULT 6
T41588
probable amino-acid permease - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000

C:Accession: T41588
R:Murphy, L.; Harris, D.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z21968
A:Accession: T41588

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-557 <MUR>
A:Cross-references: EMBL:AL031543; PIDN:CAA20834.1; GSPDB:GN00068; SPDB:SPCC74.04
A:Experimental source: strain 972h; cosmid c74
C:Genetics:
A:Gene: SPDB:SPCC74.04
A:Map position: 3
C:Superfamily: choline transport protein

Query Match 15.9%; Score 429; DB 2; Length 557;
Best Local Similarity 24.1%; Pred. No. 1.7e-23;
Matches 130; Conservative 110; Mismatches 262; Indels 38; Gaps 11;

```
QY 3 SRPGYGEKGTQRTKNTETAAAGGASELNVPLEKKQFGTITIVSLAFVICNSWAGIS 62
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 27 SNTPNEPNTSEADNEDLAALGYKEF-----QRLSLFSVFSFSLGLLPSVA 80

QY 63 GSLQALLAGGPVTLLYGILSTLYVYICIAFSLAELTSYPTAGQYHFASTILAPKSNR 122
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 81 TTLPSYIGYGTGPGLLGWLIAWVFIICIALSMAELCSAMPTSGGLYAAAVLAPEGWGP 140

QY 123 SISYVGLVSLLSWTAICSSVTMIPAQIIPALIAYSHTYSQDSWHVFLIYEGVALVLL 182
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 141 LAAMFTGSNVIQAOLVGPGSNYSTAAMLLGAVNIGNPNYEQYQLFLVSTAIQFIHF 200

QY 183 FNLFAKRNPNWHEIGFGLTIALFVISTAILARSNPK---APNSQVMTAWNNTYGTSGD 239
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 201 LASMPTKYIAKLSNVGTNLTLFLFISMLVILAMSKNHGFNETSKVSHIENITDWDPG 260

QY 240 VCFILGSTCFMFGLDAAHMAEECTDAARTVPKAVVSALIIIGFCTAPFPYTIAYLGI 299
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 261 FAILMFCGVITWMSGYDAPFHMSEETANASVNAPRGIILTAIIGMGVWQIVIAIVT 320

QY 300 TDLDSILSSAGYIPETWTQSLRSLSPATVLSGGIVNAFFALNAVOETASRLTWSFARD 359
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 321 VDOTAVTGSDSMWATYLSQCLPKRAALGILSL-TIVSFLMGQNLIASSRIAYSARD 379

QY 360 NGLVFSTLERHPWQVPMVSLFTWGI-----LATGCGCLFLGSSAFNALVNSAVVL 413
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 380 GVLPSSEWATVNPITKPIRAVFNVFGVILLFLAFAGATIG-----AVFSVTAIA 433

QY 414 QQLSFLIPIALLLYOKRDPKFLPSTRAFLV---PRGIGFLNVLAVFTVTTFVFPPL 470
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 434 AFTAPVAPVAMRVFPVKDADE--RTGPNLGFKSPKIGF---CSVSFVALMIPILCPFS 487

QY 471 TVPTAASTMYNTSAIIGVALALGVNLW-VVIARKHYQGPHELE-----DGRVVGAEFQVG 524
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 488 VKNPTPAEMNNWTCLVFGAPM-LAVLIWYAIQSKRWFKGPRINLASEGDNSTLEGVELYTG 546
```

RESULT 7
S62527
probable amino-acid permease - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 10-Dec-1999
C:Accession: T41435; T39133; S62527
R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, October 1998
A:Reference number: Z21993
A:Accession: T41435
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-544 <SEE>
A:Cross-references: EMBL:AL032824; PIDN:CAB37426.1; GSPDB:GN00068; SPDB:SPCC584.13
A:Experimental source: strain 972h; cosmid c584
R:Lye, G.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.

submitted to the EMBL Data Library, November 1995
A:Reference number: Z21830
A:Accession: T39133
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-544 <LYE>
A:Cross-references: EMBL:Z66569; NID:g1052533; PIDN:CAA91521.1; PID:g1052542; SPDB:S

C:Genetics: <SEE1>
A:Gene: SPCC584.13
A:Map position: 3
C:Genetics: <LYE2>
A:Gene: SPAC84.11
A:Map position: 1
C:Superfamily: choline transport protein

Query Match 15.6%; Score 420; DB 2; Length 544;
Best Local Similarity 24.1%; Pred. No. 7.4e-23;
Matches 120; Conservative 103; Mismatches 231; Indels 44; Gaps 11;

```
QY 39 KKQFGTITIVSLAFVICNSWAGISLQALLAGGPVTLLYGILSTLYVYICIAFSLAEL 98
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 38 KRFSAWTSCVFSFVGLLPSFASTMYTGTAGTPAMVWGLIAMVFOCVANGMAEL 97

QY 99 TSYPTAGQYHFASTILAPKSNRSISYVGLVSLLSWIAIGSSVTMIPAAQIIPALIAAY 158
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 98 CSSMPTSGGLYAAAVLAPKMGWGPFAAWLTGWSNVLQVTPGPSVAVSFAGMILTQLH 157

QY 159 SHYQSODSHVFLIYEGVAL-----VLLFNLFALKRNPWVHIEIGFGLTIALFV 207
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 158 NPNFETQNYQIFLLAAVAAMIAQGFISSMPTKVLAVFNTWGTVLNN-----LFIAIVM 209

QY 208 ISFIATILARSNKAPNS--OVMTAWSNVTGWSGVCFTLGLSTSCFPIGLDAAHMAEE 265
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 210 ITVLAVAGTKTPGPNHNVWNEFDQTDWSNGMAMLSFAGVITWMSGDVSDFHLSSE 269

QY 266 CTDAARTVPKAVVSALIIIGFCTAPFPYTIAYLIGITDLSILSSAGYIPFETMTQSLRSL 325
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 270 CSNASVAAPRAIVMTSAFGGIVGWLNLNLCIATYITVDVNAAMNDLGGPFVYVLRQVCNK 329

QY 326 FATVLSGGIVMAFFALNAVOETASRLTWSFARDNGLVFSTHLERHPRWQV---VWSL 382
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 330 TTVALTSLTVICSEFMGQGCMAASRVTSYARDGVFPFSKYLAIYDKRTKTPNVCVM-M 388

QY 383 FATMGILATCGCTFLGSSAFNALVNSAVVLQQLSELIPIALLLYOKRDPKFLPSTRAFV 442
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 389 NVVVGILC-CLLIPAGEA-AINALFSVGAIAAFVFTPIFLRVFFVKEDFEKRG----- 441

QY 443 LPRGIGFLNV---LAVVFTSVTTFVFPPL-----TVPTAASTMYNTSAIIGVALALGV 494
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 442 -PMHLGKFSKINGVAAACAFVLLMVPILCFPQFRGKDNTPDA---MNWTCVVFGGPMMLAVL 497

QY 495 LNVVHARKHYQGPHELEL 512
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 498 IMWFVSARKWFKGPRLTI 515
```

RESULT 8
S30253
GABA transport protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: GABA-specific permease; protein D1037; protein YDL210w
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: S30253; S67769; S25147
R:Andre, B.; Hein, C.; Gresson, M.; Jauniaux, J.C.
Mol. Gen. Genet. 237, 17-25, 1993
A:Title: Cloning and expression of the UGA4 gene coding for the inducible GABA-speci
A:Reference number: S30253; MUID:93204891
A:Accession: S30253
A:Molecule type: DNA
A:Residues: 1-571 <AND>
A:Cross-references: EMBL:X66472; NID:g4749; PIDN:CAA47101.1; PID:g4750
A:Note: the sequence from Fig. 5 is inconsistent with that from Fig. 3 in having 527

R:Schmidt, E.R.; Bahr, A.; Kraemer, C.; Hankeln, T.; Moeller-Rieker, S.
submitted to the Protein Sequence Database, July 1996

A:Reference number: S67756

A:Accession: S67769

A:Molecule type: DNA

A:Residues: 1-571 <SCH>

A:Cross-references: EMBL:Z74258; NID:g1431349; PIDN:CAA98788.1; PID:g1431350; GSPDB:GN00

A:Experimental source: strain 5288C

C:Genetics:

A:Gene: SGD:UGA4; MIPS:YDL210W

A:Cross-references: SGD:S0002369; MIPS:YDL210W

A:Map position: 4L

C:Superfamily: choline transport protein

C:Keywords: transmembrane protein

F:79-95/Domain: transmembrane #status predicted <TM1>

F:108-124/Domain: transmembrane #status predicted <TM2>

F:154-170/Domain: transmembrane #status predicted <TM3>

F:203-219/Domain: transmembrane #status predicted <TM4>

F:229-245/Domain: transmembrane #status predicted <TM5>

F:320-336/Domain: transmembrane #status predicted <TM6>

F:365-381/Domain: transmembrane #status predicted <TM7>

F:420-436/Domain: transmembrane #status predicted <TM8>

F:485-501/Domain: transmembrane #status predicted <TM9>

Query Match 14.2%; Score 383.5; DB 1; Length 571;
Best Local Similarity 25.8%; Pred. No. 3.4e-20;
Matches 142; Conservative 107; Mismatches 236; Indels 65; Gaps 22;

Qy 6 SCYGEKG--GTRQTTKNTETAAGGASRLNPLKQFGTTITIVSLAFVICSNWAGISG 63

Db 41 TGAGEVNYDAKSKSVNDQQLAEGYKQEL-----KRFSTLOVGIAFISGLLPSEA- 94

Qy 64 SLQIALLAGGPVTLGYLILSTLVYICIAFSLAELTSVYPTAGQYHFASILAPKSINS 123

Db 95 SVMGGGLGGPATLVWGFVAFFLLVGIITMAEHASSTIAGGLYYWYYVAPEGYKRI 154

Qy 124 ISYVGLSVLSLWSIAIGSSVTMIPAQOIPALIAAASHYVSQD-----SWHFVFLYEGVA 177

Db 155 ISFIIGCSNLALAAAGVCSIDYGLAEETAAAV-----TLTKDGNFVTSKGLYGFAGAV 209

Qy 178 LVVLFNLFALKRNPWPVHEIGFGLTIALFVIFIAILARSNPKA-----PNSQVWTAWSNY 233

Db 210 VVMCICTVASGAIARLOTLSIFANLFIIVLLFIALPTGKTRMGFGNDGDFIFGKYENL 269

Qy 234 TGWSDGVCTIL-GLSTSCFMFIGLDAAMHLAECDTAAARTVPAKVSAIIIGCTAFPPY 292

Db 270 SDWNGWQFCLAGFPAVWTIGSFDCVHQSEAKDKKSPVIGITISSIAVCWILGWLLI 329

Qy 293 IAVLYGIT-DLDSILSSA-GYIPFETWTQSLR--SLSFATVLSGCGIVMAFFALNAVOE 347

Db 330 ICLMACINPDIDSVLDSKYGALAOIIVDSLGKKWAIAPMSL-----TAFQFLMGASIT 384

Qy 348 TALSRLTSFARDNGLVFSTHLERHPRQVPMVSLFATWGLATG-CGIFLG-----SS 400

Db 385 TAVSQWAFSRDNGLPISKYIKRVDSKYVPPFA-----TLACVGSLLIIGLCLIDD 438

Qy 401 TAFNALVNSAVVLOQLSFLIPTALLLYOKRDKPFLPSTRAFLVPRIGFLVNVAVFTS 460

Db 439 AATDALFSIAGVAGNNLANSTPTVFRLTSGRD-LFRPG--PEYLGKWSPIVAWTGVAFQL 495

Qy 461 VTTVFESFP-----LTVPTASTMNTYTAIGVAL-ALGVLNWVHARKHYOGPHLELDG 514

Db 496 FIILVMPFPOHQHIT-----KSTMNY-ACVIGPGIWLAIYIKYKKYHGPATNLS 550

Qy 515 ----RWVGA 520

Db 551 DDYTEAVGAD 560

RESULT 9

T37519

probable amino acid permease - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T37519
R:Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, December 1995
A:Reference number: Z21720
A:Accession: T37519
A:Status: preliminary; translated from GB/EMBL/DBU
A:Molecule type: DNA
A:Residues: 1-550 <CON>
A:Cross-references: EMBL:Z68166; PIDN:CAA92309.1; GSPDB:GN00066; SPDB:SPAC11D3.08c
A:Experimental source: strain 972h-; cosmid c11D3
C:Genetics:
A:Gene: SPDB:SPAC11D3.08c
A:Map position: 1
C:Superfamily: choline transport protein

Query Match 13.1%; Score 351.5; DB 2; Length 550;
Best Local Similarity 23.4%; Pred. No. 6.7e-18;
Matches 129; Conservative 107; Mismatches 238; Indels 77; Gaps 16;

Qy 10 EKGGRQTTKNTETAAGGASESLN-----VPLEKKQFGTTITIVSLAFVICSNW 58

Db 5 DKESLQKQKQWENEVELGESREANQDELLMSLGYKPEPTREFSYVIFQSGFSGMGLC 64

Qy 59 AGISGSLQALLAGGPVTLGYLILSTLVYICIAFSLAELTSVYPTAGQYHFASILAPK 118

Db 65 PAMAGSLIFSMNCGG-GMYMSWIIIGICILIPVISIGELASSMPTSGGLYFWIFTLASP 123

Qy 119 SINRSTSYVGLSVLSLWSIAIGSSVTMIPAQOIPALIAAASHYTSQDSWHVFLYEGVAL 178

Db 124 SSRFTLCWCGYVSLGIATYIYASTVYSASSWQALAVIGSPSPKTYEYGIYAALLF 183

Qy 179 VVLLFNLFALKRNPWPVHEIGFGLTIALFVIFIAILARSNPKNQ--VWTASNTGW 236

Db 184 VISAMTAIPSRVIAKVINIITFOFLVSIILIIAAGSDSTTRNSGSGFICGDTNTYSCW 243

Qy 237 SD-GVCFILGSLTSCFMFIGLDAAMHLAECDTAAARTVPAKVVS----AIIIGCTAFPPY 291

Db 244 SNMGWAFILSFTTPVWVYVSGFESSAAVAEESTNAAKAAPFAMISLSGVATILGWCIVT-- 301

Qy 292 TIAVLXGITDLSL-SSAGYIPFETWTQSLR-----LSFATVLSGCGIVMAFFA 341

Db 302 TVVATWG-HDFNALIGSSLGQPVQVQVNVNNGKALGIFSLIVIALCLNCSILLIA--- 357

Qy 342 LNAVQETASRLTSFARDNGLVFSTHLERHPRQVPMVSLFATWGLATGCGIFLGSST 401

Db 358 -----ASREVFACRDGGIPGSRYL-RLITKQKVPNLAILLVLLYSLVLLILVNVVT 409

Qy 402 AFNALVNSAVVLOQLSFLIPTI-ALLLYOKRDKPFLPSTRAFLVPRIGTGL-----VNVL 454

Db 410 AISSVFNLAIIALYIAYSGPLMCRFYVKNFQ-----GVFYVQKSKPAALW 456

Qy 455 AVVFTSVTTVFESFPLTPVPTAASTMNTYTAIGVALGVLNWVHARKHYOGPHLELDG 514

Db 457 SLVWFMFMILMLLFPQYQKPNQDEMNWAIIVLGFVWFVCVV-----YYLP----- 502

Qy 515 RVYGAEFQVGP 525

Db 503 KIGGKTFFTCGP 513

RESULT 10

S38004

probable transport protein YKL174c - yeast (Saccharomyces cerevisiae)

N:Alternate names: probable transport protein YKL639

C:Species: Saccharomyces cerevisiae

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999

C:Accession: S38004; S44586; S38406

R:Vandenbol, M.; Bolle, P.A.; Dion, C.; Portetelle, D.; Hilger, F.

submitted to the Protein Sequence Database, March 1994

A:Reference number: S37976

A:Accession: S38004
A:Molecule type: DNA
A:Residues: 1-618 <VAN>
A:Cross-references: EMBL:Z28174; NID:9486305; PIDN:CAA82016.1; PID:9486306; GSPDB:GN0001
A:Experimental source: strain S288C
R:Vandenbol, M.; Bolle, P.A.; Dion, C.; Portetelle, D.; Hilger, F.
Yeast 10, 25-33, 1994
A:Title: Sequencing and analysis of a 20.5 kb DNA segment located on the left arm of yeast chromosome 1
A:Reference number: S44583
A:Accession: S44586
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-618 <VA2>
A:Cross-references: EMBL:Z26878; NID:9407503; PIDN:CAA81512.1; PID:9407507
A:Experimental source: strain S288C
C:Genetics:
A:Gene: MIPS:YKL174C
A:Map position: 1L
A:Superfamily: choline transport protein
C:Keywords: transmembrane protein

Query Match 11.98; Score 321; DB 1; Length 618;
Best Local Similarity 24.68; Pred. No. 1.2e-15;
Matches 116; Conservative 99; Mismatches 208; Indels 48; Gaps 15;

QY 37 LEKQFGTITVSLAFVICSNWAGISGSLQALLAGGPVTLTYGILISTVLVYICAF--- 93
DB 51 LDKLSRGSIVGLGLMSPVLCMTSMAITLNGPLTLMGLISG---VCLWFSLS 107
QY 94 SLAELTSYYPAGQYHPAS-ILAPKSNRSISYVCGILVLSLWTAIGSSVTMIPAQOIP 152
DB 108 SLGEIVSKFPM---ELHVGSAMLAPEKLVCSWYTGWLMIGNWTMTSTIFAGAQTLTI 164
QY 153 ALIAAYSHYTSQDSWHVF-----LIVEGVALVVLLENLFAKRNK-----WVHEIGL 201
DB 165 SLILMTNSLISEAHLFIYTVIVVTVVGLVNLKFAETINKVGVYMI----- 217
QY 202 TIALFVTSFIAFLA-----RSNPKAPNSOVMTAMSNYTCWSDG-VCFILGLSTSCPMFGL 256
DB 218 ---IYALIFIDILLVHKGFERSLKVALFHFDDNNLSYKSAFLSFIIGFOOSNFTLOGF 274
QY 257 DAAMHLEAECDAARTVPKAVSAIIIGCFAPPYTIAVLYGIDTDLSDILSAGYIP--- 313
DB 275 SMLPALADEVKVPEKIDPRGMSNAVLISAFSGVIFLIPILIPDNLFTNHNKVLPIVN 334
QY 314 -FERMTQSLRSLSPATVLSGCGIYMAFPALNAVQETASRLTWSFARDNGLVFSTHLERIH 372
DB 335 IFTKSTDSVLSPFLVLVLIILGNLL--FSGIGSI-TTSSRAVYSFSDQAIPYDKWTYVE 391
QY 373 PRWQ--VPVMSLFATWGITATCGCIFGSGTAFNALVNSAVVLOQLSFLPIALILLYQKR 430
DB 392 PDSQSKVPKNSVLSMIISYFGLLALISTAFNAFAGAAVLCSATFIPVLVLFTRR 451
QY 431 DPKELPSTRAVLPBPGIGFLVNLVAVFTSVTTVFSPPLVPTAASTMYN 481
DB 452 --RAIRSAVPKIRYK-FGWFNINISIVVLLLSMSVCLPTQVPVTFKMTNY 499

RESULT 11

T41612
amino acid permease - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: T41612

R:Murphy, L.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, May 1998

A:Reference number: Z21990
A:Accession: T41612
A:Status: preliminary;
A:Molecule type: DNA
A:Residues: 1-554 <MUR>
A:Cross-references: EMBL:AL023595; PIDN:CAA19131.1; GSPDB:GN00068; SPDB:SPCC794.03

A:Experimental source: strain 972h-; cosmid c794
C:Genetics:
A:Gene: SPDB:SPCC794.03
A:Map position: 3
C:Superfamily: choline transport protein

Query Match 11.6%; Score 311.5; DB 2; Length 554;
Best Local Similarity 25.6%; Pred. No. 5.3e-15;
Matches 121; Conservative 88; Mismatches 221; Indels 43; Gaps 17;

QY 51 AFVICSNWAGISGSLQALLAGGPVTLTYGILISTVLVYICAFSLAELTSVYPTAGQYH 110
DB 54 SFAACDCMNSIRGSFYIGLTGGPSAYWYIIAIPLIQLISAATMAEVCSALPTAGSLYF 113
QY 111 FASILAPKSNRSISYVCG--LVLSLWTAIG-SSVTMIPAQOIPALIAAYSHYTSQDS 166
DB 114 WASAAGKKYGRIGIVAWWVAVTSTFVAVNCQSTTKFIFGELPVNSGFSVSSSDVK 173
QY 167 WHV--FLIYEGVALVVLLENLFAKRNPNVHIEICGLTIALFVISTFIAI-LARSNPKAPN 223
DB 174 FRAVQNAVGEATILLVCVLLNFIPPKWFRYIFRVSVAVILLDFVLMWLPIAVSTKYGR 233
QY 224 SOVWTAMSNY-----TGWSDGVCFILGLSTSCPMFGLDAAMHLEAECDAARTVPK-- 275
DB 234 DEAFMKSTNYDLGKVNNGWS---WCLTFFCTARILVGYDAAGHVAEETKNASKTASRGM 289
QY 276 ---AVVSAIL-IGFCTAPFYTIAVLYGIDTDLSDILSAGYIPETMTQSLRSLSFATVLS 331
DB 290 FYSAFSNALISGIIIVNLYCLPPSNVMIYELIKNSQOPFVSFYAYALGRAHVFMNVV- 348
QY 332 CGGIVMAFFALNAVQETASRLTWSFARDNGLVFSTHLERIHPRMQVPVWSLFAWGLAT 391
DB 349 --GILCMIFDTSLSIVASSRLVFAVARDGVLPFSGWLRKVDVSHQO-PTNAVTFEFLISAA 405
QY 392 CGCIFGSGTAFNALVNSAVVLOQLSF-LIPIALILLYQKRQ-PKFLPS----TRAFVLP 445
DB 406 LLCSNLPSSAVAFSTLSLSAAVPTIMAYAAVAFGRFLFSRNDPFPKSEWSGLSKSPFOL-- 463
QY 446 GIGFLVNLVAVFTSVTTVFSPPLVPTAASTMYNTSIIIGVALGVLNWNV 498
DB 464 -ITFLMN----LFTAV--ILFS-PKAYPVTKGNFNPVAPVIFGATITFGLISWL 508

RESULT 12
T39195
probable amino acid permease - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: T39195
R:Wedler, H.; Duesterhoeft, A.; Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, October 1999

A:Reference number: Z21834

A:Accession: T39195
A:Status: preliminary;
A:Molecule type: DNA
A:Residues: 1-591 <WED>
A:Cross-references: EMBL:AL121764; PIDN:CAB57428.1; GSPDB:GN00066; SPDB:SPAC9.10
A:Experimental source: strain 972h-; cosmid c9

C:Genetics:

A:Gene: SPDB:SPAC9.10

A:Map position: 1

C:Superfamily: choline transport protein

Query Match

Best Local Similarity 10.2%; Score 274; DB 2; Length 591;
Matches 124; Conservative 83; Mismatches 212; Indels 118; Gaps 21;

QY 42 FGTITIVS---LAFVICSNWAGISGSLQALLAGGPVTLTYGILISTVLVYICAFSLAEL 98
DB 87 FASLDVWSGVRLTF-----SW-GIS-----FGGPAVWSAMLVTCFCSIVTAACLAEL 133


```
Db 109 GWDILEYMLAVSAVSGWGYFQSFSLGSLGHLFPV-ALTAAPGAVKGTFTLNLPAFVI 167
QY 173 YEGVALVLLFNLFALKRNPWHIEGFGTIALFVLSFTAILARSNPKAPNSOVWTA--- 229
Db 168 V--MAITYLLY--LGKESKRNNMWMUKI--LVLLFTFAVA-----VYVKPHN 212
QY 230 WSNYTGWSGVCFILGLSTSCFMFIGLDAAMHLAEECTDAARTVPKAVVSATIIIGCFYAF 289
Db 213 WOPFPMFGGGVP--SAAALVFFAFIGFDVAVSSAAETKPAKDLPKGIIFSLV--CTIL 269
QY 290 PVTIAYLGIITDLSILSSAGYIPF-----ETWQTSRLSRSFATVLSGCGVWAF 339
Db 270 YVTVSAT-----MTGVIPFAGVAGVHPSVLQSQAGNWWAGIIDIGAVLGWT 318
QY 340 FALNAVQETASRLTWSFARDNGLVSTHLERIHPRWQVPVMSLFATWGLATCGCIFLGS 399
Db 319 TVMLVNLVYQTRVPMFMSRD--GLVFGS--LSKVHPKHKTP---YVATW-----FFGT 364
QY 400 STAFNALVNSAVVLOQLS-----FLIPIALLLYOKRDKPFLPSTRAFLVLPRTGIG 448
Db 365 ---MSALLGSLVPLDELAKLVNIGTSLAFVLISVAVIYLRKKQPD-LP--RAFKCP-GVP 417
QY 449 FLVNVLAIVFTSVTVFFSEPLTPTVTAASTMNYTSIIAGIYALALGVLNKNVHARKH 504
Db 418 -VIPGLAILF---CLFLIILNGWTIVRFL-----VWLLIGVIYFLYSRKH 460
```

RESULT 15

```
C64878
probable amino acid permease ycjJ - Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 17-Mar-2000
C:Accession: C64878
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: C64878
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-479 <BLAT>
A:Cross-references: GB:AE000027; GB:U00096; MID:g1787543; PIDN:AAC74378.1; PID:g1787553;
A:Experimental source: strain K-12, substrain MGI655
C:Genetics:
A:Gene: ymjJ
C:Superfamily: arginine permease
C:Keywords: amino acid transport; transmembrane protein
F:73-89/Domain: transmembrane #status predicted <TM1>
F:130-146/Domain: transmembrane #status predicted <TM2>
F:150-166/Domain: transmembrane #status predicted <TM3>
F:181-197/Domain: transmembrane #status predicted <TM4>
F:222-238/Domain: transmembrane #status predicted <TM5>
F:260-276/Domain: transmembrane #status predicted <TM6>
F:362-378/Domain: transmembrane #status predicted <TM7>
F:383-399/Domain: transmembrane #status predicted <TM8>
F:419-435/Domain: transmembrane #status predicted <TM9>
```

```
Query Match 8.4% Score 226.5; DB 2; Length 479;
Best Local Similarity 22.1%; Pred. No. 6.3e-09;
Matches 100; Conservative 70; Mismatches 139; Indels 143; Gaps 18;
```

```
QY 25 AAGGASESINVLPL-----EKKQFGTITIVSLAFV-----ICNSWAGISGSLQLA 68
Db 13 AGEAAIMAINPLNTAAOPGTRLRKSLKLMQVMMGLAYLTPMTVFTFGIVSG----- 67
QY 69 LIAGGPVTLXGILISTLVYICIAFLAELTSVYPTAGQYHFASILAPKSNIRNSISYVC 128
Db 68 -ISDGHVPASVLLAGLVLETAISY--GKLVQRPFAGSAVYTAQ-----KSINPHVGPVW 120
QY 129 GLVSLLSWIAIGSSVTMIPAAQIIPALIAAYSHTYSQDSWHVFLIYEGVALVLLLENLAL 188
```

```
Db 121 GWSLDDVL-----FLPMINVLLAKIYLSALFP 148
QY 189 KRNPAVHEIGFGLTIALFVLSFTAILARSNPKAPN----- 223
Db 149 EVPPWV-----WVTFEVALITAAALKSVNLVANFNTLFLVLSQISIMVVFILVV 197
QY 224 -----SOVWTANSVYTGWSDG---VCFILGLSTSCFMFIGLDAAMHLAEECTDAARTVP 274
Db 198 OGLHKGEVGTVMSLQPPFISENAHLIPIITGATIVCFPSFLGFDVAVTTLSEETPDAARVIP 257
QY 275 KAVSAILIIGFCTAPPYTIAYLVYGITDLSILSSAGYIP-----FETWQTSRLSLS----- 325
Db 258 KAI-----FLTAVYGVGVIFFAAAFMQLFFPDISRFRKDPDAALPEIALLYGG 304
QY 326 --FATVLSGCGIVMAFFALNAVOETASRLTWSFARDNGLVSTHLER-----IHPRWQVPV 379
Db 305 KLFQSIPLCTTFTVNTLASGLASHASVRLYYVMGRDN--VFP--ERVFGYVHPKWRTPA 359
QY 380 WSLFATWGLATCGCIF-LGSSTA---FNALV 407
Db 360 LNVIMV-GIVALSALEFFDLVTATALINFCALV 390
```

Search completed: April 27, 2002, 07:48:23

Job time: 6730 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:58:37 ; Search time 24.91 Seconds
(without alignments)
147.422 Million cell updates/sec

Title: US-09-882-694A-8
Perfect score: 2692
Sequence: 1 MDSRSGYGERGGTRQTKN.....QGFHLELDGRVVGAEFQVCP 525

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 40831 seqs, 6994805 residues

Total number of hits satisfying chosen parameters: 40831

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	171.5	6.4	714	1	PCT-US02-09944-473
2	159	5.9	585	5	US-09-573-655B-1554
3	156	5.8	519	5	US-09-895-913A-118
4	133.5	5.0	533	5	US-09-540-209B-6591
5	123	4.6	832	5	US-09-573-655B-1490
6	116.5	4.3	562	5	US-09-540-209B-5760
7	112	4.2	459	5	US-09-540-209B-6435
8	106.5	4.0	466	1	PCT-US02-09921-675
9	106	3.9	483	5	US-09-573-655B-2098
10	105	3.9	456	1	PCT-US02-09962-17
11	103.5	3.8	867	5	US-09-573-655B-364
12	102	3.8	1089	6	US-10-006-063A-102
13	102	3.8	1089	6	US-10-006-117A-102
14	102	3.8	1089	6	US-10-006-130A-102
15	102	3.8	1089	6	US-10-006-172A-102
16	102	3.8	1089	6	US-10-006-768A-102
17	102	3.8	1089	6	US-10-017-527A-102
18	102	3.8	1089	6	US-10-017-610A-102
19	100.5	3.7	429	5	US-09-540-209B-10419
20	100.5	3.7	450	5	US-09-540-209B-5857
21	98.5	3.7	488	6	US-10-109-310-15
22	98.5	3.7	530	6	US-10-107-431-275
23	97.5	3.6	466	5	US-09-540-209B-6619
24	97	3.6	515	5	US-09-540-209B-10338
25	96	3.6	340	6	US-10-108-605-291
26	96	3.6	474	5	US-09-540-209B-6787

27	95.5	3.5	426	5	US-09-540-209B-5807	Sequence 5807, Ap
28	95.5	3.5	584	1	PCT-US02-09921-601	Sequence 601, App
29	95	3.5	445	6	US-10-107-431-91	Sequence 91, Appl
30	94	3.5	353	1	PCT-US02-05625-63	Sequence 63, Appl
31	93	3.5	498	5	US-09-540-209B-5414	Sequence 5414, Ap
32	93	3.5	836	5	US-09-540-209B-6015	Sequence 6015, Ap
33	92.5	3.4	442	1	PCT-US01-25881-26	Sequence 26, Appl
34	92.5	3.4	626	1	PCT-US02-07826-95	Sequence 95, Appl
35	92.5	3.4	626	6	US-10-097-340-95	Sequence 95, Appl
36	92	3.4	443	5	US-09-540-209B-9709	Sequence 9709, Ap
37	91	3.4	390	5	US-09-895-913A-286	Sequence 286, App
38	91	3.4	536	1	PCT-US02-09962-16	Sequence 16, Appl
39	90.5	3.4	430	5	US-09-540-209B-5690	Sequence 5690, Ap
40	90.5	3.4	462	6	US-10-108-605-345	Sequence 345, App
41	90.5	3.4	525	5	US-09-972-211-104	Sequence 104, App
42	89.5	3.3	521	5	US-09-972-211-106	Sequence 106, App
43	89.5	3.3	630	5	US-09-540-209B-7741	Sequence 7741, Ap
44	89.5	3.3	637	5	US-09-540-209B-5999	Sequence 5999, Ap
45	89	3.3	439	5	US-09-540-209B-5802	Sequence 5802, Ap

ALIGNMENTS

RESULT 1
PCT-US02-09944-473
; Sequence 473, Application PC/TUS0209944
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: DAFFO, Abel
; APPLICANT: JONES, Anissa L.
; APPLICANT: TRAN, Alanna-Phung B.
; APPLICANT: DAHL, Christopher R.
; APPLICANT: GIETZEN, Darryl
; APPLICANT: CHINN, Joyce
; APPLICANT: DUFOUR, Gerard E.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: YU, Jimmy Y.
; APPLICANT: TUASON, Olivia
; APPLICANT: YAP, Pierre E.
; APPLICANT: AMSHEY, Stefan R.
; APPLICANT: DAUGHERTY, Sean C.
; APPLICANT: DAM, Tam C.
; APPLICANT: LIU, Tommy F.
; APPLICANT: NGUYEN, Duy-Viet An
; APPLICANT: KLEEFELD, Yael
; APPLICANT: GERSTIN, JR., Edward H.
; APPLICANT: PERALTA, Careyna H.
; APPLICANT: DAVID, Marie H.
; APPLICANT: LEWIS, Samantha A.
; APPLICANT: CHEN, Alice J.
; APPLICANT: PANZER, Scott R.
; APPLICANT: HARRIS, Bernard
; APPLICANT: FLORES, Vincent
; APPLICANT: MARWAHA, Rakesh
; APPLICANT: LO, Audrey
; APPLICANT: LAN, Ruth Y.
; APPLICANT: URASHKA, Michael E.
; TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT
; FILE REFERENCE: PT-1231 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09944
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/279,619; 60/280,067; 60/279,619; 60/280,068; 60/291,280; 60/291,818
; 60/291,829; 60/299,428; 60/300,001; 60/299,776
; PRIOR FILING DATE: 2001-03-29; 2001-03-28; 2001-03-29; 2001-05-16; 2001-05-17;
; 2001-05-17; 2001-06-19; 2001-06-20; 2001-06-20
; NUMBER OF SEQ ID NOS: 792
; SOFTWARE: PERL Program
; SEQ ID NO 473
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:

Db 182 IALVFL--NFFSVKIFA--EGEFSEFLIKVLAVIAFIGA-----IGIIYQIYSHGF 232
Qy 236 -----WSDGVCFIL--GLSTSCFMFIGLDAAMHAEECTDAARTVPRKAVSAI 281
Db 233 SIFDNHFGDKGFPNGSAAVFSAAMLAVIAFAGTGVIGVAVGETKNASEVMPKAIKATL 292
Qy 282 --IIGCTAPPYIAVLYGIDTDLSDISSAGYIPFETMTQSLRSLF-----AVVLS 331
Db 293 WRIVFFLGSVVISFLPND-----SSITQSPFVSFLERI--NLPIGIMGIPYVADIMN 346
Qy 332 CGGIVMAFFALNAVQETASRLTWSFARDNGLVFTSHLERIHPRQVPMVMSLFAFWGILAT 391
Db 347 AVIITAMFSTANSGLYGASRMIVGLSKQK--MEFKVFSQLN--RQCTTYAMFFS----- 397
Qy 392 CGCIFLGSSTAFNALVNSAVVLOOLSF-----LPIALLLYOKRDPKFLPSTRAFVLRGIGFL 450
Db 398 -----LSPSLGILLVQIYAKENV-----VEAL 419
Qy 451 -VNVLA---VFTSVTVTFSPFL-----TVPTAASTMNYTSALIGVALALGVLN 496
Db 420 INVISFTVILVMSVSQYSFRKQYLKAGHSLEDLPYKAPFLPFLQILIGITGCAIGVI- 478
Qy 497 WYVHARKHYOGPHLELDGRV 516
Db 479 -----GSAMDKDQRI 488

RESULT 4

US-09-540-209B-6591
; Sequence 6591, Application US/09540209B
; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FR
; FILE REFERENCE: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540, 209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 6591
; LENGTH: 533
; TYPE: PRT
; ORGANISM: B. fragilis
US-09-540-209B-6591

Query Match 5.0%; Score 133.5; DB 5; Length 533;
Best Local Similarity 20.7%; Pred. No. 0.0018;
Matches 106; Conservative 105; Mismatches 201; Indels 101; Gaps 28;

Qy 40 KQFGTITIVSLAFVICHWSAGISGLQIALLAGGPTVLLYGI-----LISTLVYICIAF 93
Db 6 KOAVKLGVFILA--IMNVTVVS-----LRGLPAEAVYGMSSAFYVLFIAIVFL-IPT 55
Qy 94 SL--AELTSVY-PTAGQYHFASILAPKSNRSISYVGLVLSLSW-----TAIGS-SVTM 145
Db 56 SLVAELAAAMFQDKQGGVFRWVGGEAYKKGLGFLAIWQWIESTI-WYPTVLTFGAVSTAF 114
Qy 146 IPAQOIPALIAAAYSHTSYQSDSWHVFLEVEGVALVFLFNLFALKNRPWVHE-----IG 198
Db 115 I-----GMNDTHDWTLASNKYYTILAV---VLIYWLATFISLKGCMGVKVGKIGMVG 165
Qy 199 FGLTALFVIFAILA---RSNPKAPNSQVWTAMSNYTGSDGVCFILGISTCFMF-I 254
Db 166 TIIPAAALLIILGIVLASGGHSLDF-HSSFFPDITNF-----DNVV-----LAASIFLEYA 216
Qy 255 GLDRAHMLAECTDAARTVPRKAV-----VSAIIIGCTAFPTYIAVLYGIDTDLSDILSSA 309
Db 217 GWEMGGIHVKDMQNPKNYPRKAVFIGALITVILVLTG---FSLGIITIPAKDIS--LQOS 271
Qy 310 GYIPFETMTQSLRSLSFATVLSGCGIVMAFFALNAVQETASRLTWSFARDNGLVFSTHLE 369
Db 272 LLVGFNDYFRYIHASWLSPIA---IALAFGVLAGV-----LTWAGVPSKGIFAVGKAG 322

Qy 370 RIHPRWQ-----VPWMSLFATWGIATCGCICFL---GSSTAFNALVNSAVVLOOLSELI 420
Db 323 YMPPEFQKTNKLGVOKNILFVQGGAVTVLSLLFLFVWMPVSQSFYOILSQTIVILYVWYLL 382
Qy 421 PIALLLYOKRDPKFLPSTRAFVLP-----GIGELVNVLAVVFTSVTVTFSPFL 470
Db 383 MFSGAIYLYRNK--KANRPFRIGKKGKGLMWTGVLGSLGSLAFLS-----FIPPS 434
Qy 471 TVPTAASTMNYTSALIGVALALGVLNWVHARK 503
Db 435 QISTGNTWFSVLIIG-ALVVVIAPFIYAAR 466

RESULT 5

US-09-573-655B-1490
; Sequence 1490, Application US/09573655B
; GENERAL INFORMATION:

; APPLICANT: SOLOVIEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptid
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptid
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573, 655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1490
; LENGTH: 832
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-1490

Query Match 4.6%; Score 123; DB 5; Length 832;
Best Local Similarity 21.0%; Pred. No. 0.019;
Matches 106; Conservative 79; Mismatches 178; Indels 142; Gaps 29;

Qy 26 AGGASESLNVPLEKKQFGQITIVSLAFVICHWSAGISGLQIALLAGGPT-----VTL 77
Db 2 ADPAAESIDA--SSSRFGR-----VVCYQNTAHGNSNTIS--AAAPFTQLSVANL 49
Qy 78 LYGILITLVYICIAFSLAELTSVYPTAGQYHFASILAPKSNRSISYVGLVLSLSWI 137
Db 50 TYRILYFLKPLCLPFFVAQI-----LCGL-LFSPT 79
Qy 138 ATGSSVTMIPAOQOIPALIAAAYSHTSYQSDSWHVFLEVEGVALVFLFNLFALKNRPWVHEI 197
Db 80 VLGNEVVL-----KLIFPKYT-----MLLETFANLALVYNVFL----- 115
Qy 198 GFLTALFVIS-----FTAILARSNPKAPNSQVWTAMSNYTGSDGV---CFILGLSTS 249
Db 116 GLGLDRLMTIKIDIKPVIIAIVGELLAALAGAGLYLPSN--GEADKILACGMWSTAFG 173
Qy 250 CFMTIGLDAAMHLAE---ECTDAARTVPRKAVSAIIIGCTAFPTYIAVLYGIDTDLSDIL 305
Db 174 CTNF--PDARILADLKLRTDMGHT---AMCAAVTDLCT---WLLFEGM---AIF 220
Qy 307 SSAGYIPFETMTQSLRSLSFATVLSGCGIVMAFFALNAVQETASRLTWSFAR--DNGLVF 364
Db 221 SKSG-VRNEMLPYSLAS-TIAFVLLC-----YFVIQ-----PGVAIENNTVEGQVG 266
Qy 365 STHLERIHPRWQ-----VPWMSLFA-TWGIATCGCIFLGSSTAFNALVNSAVVLOOLSEL 419
Db 267 DTHV-----WYTLAGVVICSLITEVCGVHSITGAFGLGISIPHDHIIIRKMEIEKLHDFL 320
Qy 420 IFIALLLYQ-----KRDPKFLPSTRA---FVLPRGIGFLVNVLAVVFTSVTVTFSPFLT 471
Db 321 SCMLPLFYIICGLRADICGMNPTVSGVMMAVVTASVMVKILSTWECF---IFLRIPLR 377
Qy 472 VPTAASTMNYTSALIGVALALGVLN 496
Db 378 DGLAIGALMNTKG-----TWALVILN 398

RESULT 6

US-09-540-209B-5760
; Sequence 5760, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FR
; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540.209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 5760
; LENGTH: 562
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-5760

Query Match 4.3%; Score 116.5; DB 5; Length 562;
Best Local Similarity 20.7%; Pred. No. 0.038;
Matches 94; Conservative 79; Mismatches 173; Indels 109; Gaps 25;

QY	77	LLYGILSTLVY-TCIAFSLAELTSVYTAGGQYHFASILAPKSNRSISYVCG-LVSLL	134
DB	29	LMAGIIVGIVPALPAIAFGIA--SGVSPKG-----IIIT-----AIIAGFIISLL	71
QY	135	--SWIAIGSSVTMIPAAQOIPALIAAASHYTSQDSWHVFLIYEGVALVVL-LFNLFAL-KLR	190
DB	72	GGSKVQIQQG-----PTGAFIVIIYGIIOYGEAGLIVATLMAGILLILLGVFKLGAIIKF	126
QY	191	NPWVHEIGF--GLTIALFV--ISFIALARNPKAPNSQV--WTAMSNY-----TGMSDQVC	241
DB	127	IPYPIIVGFTSGIAVTFTTQIADIFGLNFGGKVPQDFIGKMMIYFRHFDVNNWNAV	186
QY	242	FILGLSTSCFMFIGLDAAMHLEECTDAARTVPRKAVSAIIIGCTAPPYIAVLGITD	301
DB	187	SILSI-----IIIIITRF-----SKKIPGSLIAIIVV---TIGVYLVKTYAGDS	229
QY	302	LDSTLSSAGYIPFETMTQSLRSLSFATVLSGGIVMAFFALNAVOETASRLTWSFARDNG	361
DB	230	IDTI-GRFTIKSELPEAAIPTLNWEAIKDLFPVAITIAVLGAIESLSATV-----ADG	283
QY	362	LVFSTH--LRIHPRQVQVWSLF----ATWGILATCGCITFLGSSATFALNVSADVLO	414
DB	284	VTGDKHDSNTIELIAQGTANLITPLFGGIPATGAIARTMTNINNGKTPVAGIIHAIIVLL	343
QY	415	QLSLFLIPALLLYOKRDKPFLPSTRAFLPRGIGLVNVAVVF-----	458
DB	344	ILFLMPLA-----QYIP-----MACLAGVLIVSYNNMSEWTFKALLKNP	384
QY	459	-TSVTTVFFSFPFLVPTAASTMNTSAI-IGVALA	491
DB	385	KSDVTLLITLITFLI-----IFDLTIAIEVGLVIA	414

RESULT 7

US-09-540-209B-6435
; Sequence 6435, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FR
; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540.209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 6435
; LENGTH: 459
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-6435

Query Match 4.2%; Score 112; DB 5; Length 459;
Best Local Similarity 21.6%; Pred. No. 0.067;
Matches 117; Conservative 72; Mismatches 173; Indels 180; Gaps 27;

QY	33	LNVPLEKKQFGTITIVSLAFVICSNWAGIS-GSLOLALLAG--GPVTLLYGIL-----I	83
DB	14	LSPLEKTKKK--DMILQLAFVLTIIIGARLGGIGLVGMGGVGLIILTFAPGLQPTAPPI	70
QY	84	STLVYICIAFSLAELTSVYTAGGQYHFASILAPKSNRSISYVCGL-----	130
DB	71	DYMLMIAAVISAA---SCHQAAGGLDYMVK-LAEKLLRKNPSHVITLSPIVYTLFTFVAG	126
QY	131	-----VSLLSWIAIGSSVTMIPAAQO---IPALIAAASHYTSQDSWHVFLIYEGVALVLL	182
DB	127	TGHVAYSVLPVIAEVATETKIRPERPLGIAVIAASQAATASPIS-----AATVAL	176
QY	183	FNLFALKRNPWVHEIGFGLTIALFVLSFTAILARSNPKAPNSQVWTANSYTGSDGVCF	242
DB	177	LGLLA-----GFDITLFDILKITIPAT-----	198
QY	243	ILGLSTSCFMFIGLDAAMHLEECTDAARTVPRK-----AVVSAI	281
DB	199	IIG-----VLGALFSMKVGRKELVDDEYQKRLAEGYFNKKIEIKDVHRRNAMISVL	252
QY	282	IIGCTAPPYTTIAVLGYITDLSILSSAGYIPETMTQSLRSLSFATVLSGGIVMAFFA	341
DB	253	IFILATAP-----IVPFGSFD---GMRPTFLIDGETVTLCMSAIIIEIVMLSAALILLITK	305
QY	342	LNVAQETASRLTWSFARDNGLVFSTHLERHPRQVQVWSLFATWGILATCGCICFL-GSS	400
DB	306	TGDKAT-----QGSVFPAGMQ-----AVIAIFGI-AWNGDTFLOGNM	342
QY	401	TAFNALVNSAVVQLQSLFLIPTAL-----LLYOKRDKPFLPSTRAFLVPRGI--GFLVNV	453
DB	343	GOLTESIEGLV--RQNPWLFGLIALFMSILLYSQ-----AATVRLMPLGIALGISPYM	394
QY	454	LAVVTSVTVTFSSFPFLVPTAASTMNTY---SAILG-----VALALG	493
DB	395	LIAMPVAVNGYFF-IP-NYPTVVAAINFORTGTTKIGKYVLNHSFMMPLGLISTVVAIALG	452
QY	494	VL	495
DB	453	LL	454

RESULT 8
PCT-US02-09921-675
; Sequence 675, Application PC/TUS0209921
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: DAFFO, Abel
; APPLICANT: JONES, Anissa L.
; APPLICANT: TRAN, Alanna-Phung B.
; APPLICANT: DAHL, Christopher R.
; APPLICANT: GIETZEN, Darryl
; APPLICANT: CHINN, Joyce
; APPLICANT: DUFOUR, Gerard E.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: YU, Jimmy Y.
; APPLICANT: TUASON, Olivia
; APPLICANT: YAP, Pierre E.
; APPLICANT: AMSHEY, Stefan R.
; APPLICANT: DAUGHERTY, Sean C.
; APPLICANT: DAM, Tam C.
; APPLICANT: LIU, Tommy F.
; APPLICANT: NGUYEN, Duy-Viet An
; APPLICANT: KLEEFELD, Yael
; APPLICANT: GERSTEIN JR., Edward H.
; APPLICANT: PERALTA, Careyna H.
; APPLICANT: DAVID, Marie H.
; APPLICANT: LEWIS, Samantha A.
; APPLICANT: CHEN, Alice J.
; APPLICANT: PANZER, Scott R.


```

RESULT 12
US-10-006-063A-102
: Sequence 102, Application US/10006063A
: GENERAL INFORMATION:
: APPLICANT: Baker, Kevin P.
: APPLICANT: Botstein, David
: APPLICANT: Desnoyers, Luc
: APPLICANT: Eaton, Dan I.
: APPLICANT: Ferrara, Napoleone
: APPLICANT: Fong, Sherman
: APPLICANT: Gao, Wei-Qiang
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Grimaldi, Christopher J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Hillan, Kenneth J.
: APPLICANT: Pan, James
: APPLICANT: Paoni, Nicholas F.
: TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
:   ACIDS Encoding the Same
: FILE REFERENCE: P2830P1C3
: CURRENT APPLICATION NUMBER: US/10/006,063A
: CURRENT FILING DATE: 2002-03-15
: Prior Application removed - See File Wrapper or Palm
: NUMBER OF SEQ ID NOS: 477
: SEQ ID NO 102
: LENGTH: 1089
: TYPE: prt
: ORGANISM: Homo sapiens
US-10-006-063A-102

```

Query Match	3.8%	Score 102;	DB 6;	Length 1089;
Best Local Similarity	21.0%	Pred. No. 1;		
Matches 99;	Conservative	54;	Mismatches	177;
			Indels	142;
			Gaps	26;

[illegible]

QY 376 -----QVPV-----NSLFATWGLATCGCIFLGSSAFNAL 406
Db 864 LQSFLLHLAAGIVTTPGFTVPQAVSAWALMATOTFTYSTGHQVPVPAI 915

RESULT 15

US-10-006-172A-102
Sequence 102, Application US/10006172A

GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2830P1C11
CURRENT APPLICATION NUMBER: US/10/006,172A
CURRENT FILING DATE: 2002-03-19
PRIOR APPLICATION NUMBER: 60/098716
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098723
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098749
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098750
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098803
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098821
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098843
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/099536
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099596
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099598
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099602
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099642
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099741
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099754
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099763
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099792
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099808
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099812
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099815
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099816
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/100385
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100388
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100390
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100584
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100627
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100661
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100662
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100664
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100683
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100684
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100710
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100711
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100848
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100849
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100919
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100930
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/101014
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101068
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101071
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101279
PRIOR FILING DATE: 1998-09-22
PRIOR APPLICATION NUMBER: 60/101471
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101472
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101474
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101475
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101476
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101477
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101479
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101738
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101741
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101743
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101915
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101916
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/102207
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102240
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102307
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102330
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102331
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102484
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102487

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:55:20 ; Search time 389,72 Seconds

(without alignments)
474.160 Million cell updates/sec

Title: US-09-882-694A-8

Perfect score: 2692

Sequence: 1 MDSRPSGYGKGGTROTTRKN.....QGPHELDGRVGAERQVGP 525

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_Main:*

- 1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep.*
- 2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
- 3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
- 4: /cgn2_6/ptodata/2/paa/US080_COMB.pep.*
- 5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
- 6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
- 7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
- 8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
- 9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
- 10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
- 11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
- 12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
- 13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
- 14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
- 15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
- 16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
- 17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
- 18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
- 19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
- 20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
- 21: /cgn2_6/ptodata/2/paa/US097_COMB.pep.*
- 22: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
- 23: /cgn2_6/ptodata/2/paa/US099_COMB.pep.*
- 24: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
- 25: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
- 26: /cgn2_6/ptodata/2/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2692	100.0	525	17	US-09-351-224E-8
2	2692	100.0	525	20	US-09-677-488A-8
3	2692	100.0	525	20	US-09-677-682A-8
4	2692	100.0	525	20	US-09-677-682B-8
5	2692	100.0	525	22	US-09-882-694-8
6	2692	100.0	525	22	US-09-882-694A-8
7	2655	98.6	525	17	US-09-351-224-8

Query Match 100.0%; Score 2692; DB 17; Length 525;
Best Local Similarity 100.0%; Pred. No. 3.8e-244;
Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDSRPSGYGKGGTROTTRKNETAAAGAGSESLNVPLEKKQFGTITVSLAFVTCNSWAG 60

Db 1 MDSRPSGYGKGGTROTTRKNETAAAGAGSESLNVPLEKKQFGTITVSLAFVTCNSWAG 60

8	2655	98.6	525	17	US-09-351-823-8	Sequence 8, Appli
9	2655	98.6	525	20	US-09-677-488-8	Sequence 8, Appli
10	2655	98.6	525	20	US-09-677-682-8	Sequence 8, Appli
11	561.5	20.9	585	16	US-09-248-796-20096	Sequence 20096, A
12	561.5	20.9	585	26	US-60-096-409-20096	Sequence 20096, A
13	552	20.5	503	1	PCT-US02-03987-15766	Sequence 15766, A
14	552	20.5	503	24	US-10-032-585-7822	Sequence 7822, Ap
15	552	20.5	503	24	US-10-072-851-15766	Sequence 15766, A
16	552	20.5	503	26	US-60-314-050-7822	Sequence 7822, Ap
17	544.5	20.2	450	16	US-09-248-796-20095	Sequence 20095, A
18	544.5	20.2	450	26	US-60-096-409-20095	Sequence 20095, A
19	519	19.3	411	18	US-09-417-507-33583	Sequence 33583, A
20	451	16.8	505	21	US-09-708-427-19713	Sequence 19713, A
21	451	16.8	517	21	US-09-708-427-19712	Sequence 19712, A
22	449.5	16.7	504	20	US-09-621-900A-72	Sequence 72, Appl
23	449.5	16.7	516	20	US-09-621-900A-71	Sequence 71, Appl
24	449.5	16.7	528	20	US-09-621-900A-70	Sequence 70, Appl
25	444.5	16.5	352	18	US-09-417-507-39634	Sequence 39634, A
26	400	14.9	438	21	US-09-708-427-19714	Sequence 19714, A
27	383.5	14.2	571	19	US-09-538-092-120	Sequence 120, App
28	363.5	13.5	533	16	US-09-248-796-20066	Sequence 20066, A
29	363.5	13.5	533	26	US-60-096-409-20066	Sequence 20066, A
30	349.5	13.0	381	16	US-09-248-796-20097	Sequence 20097, A
31	349.5	13.0	381	26	US-60-096-409-20097	Sequence 20097, A
32	343	12.7	426	26	US-60-324-109-17656	Sequence 17656, A
33	334	12.4	575	24	US-10-032-585-7921	Sequence 7921, Ap
34	333	12.4	268	20	US-09-675-784A-7032	Sequence 7032, Ap
35	332	12.3	475	16	US-09-248-796-20067	Sequence 20067, A
36	332	12.3	475	26	US-60-096-409-20067	Sequence 20067, A
37	292.5	10.9	202	18	US-09-417-507-24497	Sequence 24497, A
38	279.5	10.4	291	18	US-09-417-507-32921	Sequence 32921, A
39	268.5	10.0	182	18	US-09-417-507-34508	Sequence 34508, A
40	268.5	10.0	248	18	US-09-417-507-24006	Sequence 24006, A
41	268.5	10.0	283	26	US-60-324-109-25312	Sequence 25312, A
42	264.5	9.8	484	16	US-09-248-796-20357	Sequence 20357, A
43	264.5	9.8	484	26	US-60-096-409-20357	Sequence 20357, A
44	260	9.7	397	26	US-60-324-109-20988	Sequence 20988, A
45	255.5	9.5	287	18	US-09-417-507-37631	Sequence 37631, A

ALIGNMENTS

RESULT 1
US-09-351-224E-8
; Sequence 8, Application US/09351224E
; GENERAL INFORMATION:
; APPLICANT: Davick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/351,224E
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-351-224E-8

QY	61	ISGSLOALLAGGPTVLLYGLISTLVYICIAFSLAELTSVYPTAGQYHFASILAPKSI	120
Db	61	ISGSLOALLAGGPTVLLYGLISTLVYICIAFSLAELTSVYPTAGQYHFASILAPKSI	120
QY	121	NRSISYVCGVLSLWSIAIGSSVTMIPAQQIPALIAAASHYTSQDSWHVFLIYEGVALVY	180
Db	121	NRSISYVCGVLSLWSIAIGSSVTMIPAQQIPALIAAASHYTSQDSWHVFLIYEGVALVY	180
QY	181	LFNLFLKRNPNWVHEIGFGLTIALFVIFSIATLARSNPKAPNSQVWTWSFARDN	240
Db	181	LFNLFLKRNPNWVHEIGFGLTIALFVIFSIATLARSNPKAPNSQVWTWSFARDN	240
QY	241	CFILGLSTSCFMFICGLDAAMHIAEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGIT	300
Db	241	CFILGLSTSCFMFICGLDAAMHIAEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGIT	300
QY	301	DLDSILSSAGYIPFETMTQSLRSLSFATVLSGCGIVMAFFALNAVQETASRLTWSFARDN	360
Db	301	DLDSILSSAGYIPFETMTQSLRSLSFATVLSGCGIVMAFFALNAVQETASRLTWSFARDN	360
QY	361	GLVFSTHLERHPRWQVPVWSLFAFWGILATCGCIFLGSSTAFNALVNSAVVLOQLSFLI	420
Db	361	GLVFSTHLERHPRWQVPVWSLFAFWGILATCGCIFLGSSTAFNALVNSAVVLOQLSFLI	420
QY	421	PIALLLYOKRDPKFLPSTRAFLVPRGIGFLVNLAVVTSVTVFFSFPPLTVPTAASTMN	480
Db	421	PIALLLYOKRDPKFLPSTRAFLVPRGIGFLVNLAVVTSVTVFFSFPPLTVPTAASTMN	480
QY	481	YTSIIIGVALALGVNLVNVVHARKHYQGPHLELDGRVVGAEFQVGP	525
Db	481	YTSIIIGVALALGVNLVNVVHARKHYQGPHLELDGRVVGAEFQVGP	525
RESULT 2			
US-09-677-488A-8			
; Sequence 8, Application US/09677488A			
; GENERAL INFORMATION:			
; APPLICANT: Duvick, Jon			
; APPLICANT: Maddox, Joyce			
; APPLICANT: Gilliam, Jacob			
; APPLICANT: Folkerts, Otto			
; APPLICANT: Crasta, Oswald R.			
; TITLE OF INVENTION: Compositions and Methods for Fumonisin			
; TITLE OF INVENTION: Detoxification			
; FILE REFERENCE: 35718/204100			
; CURRENT APPLICATION NUMBER: US/09/677,488A			
; PRIOR FILING DATE: 2000-10-02			
; PRIOR FILING DATE: 1999-07-12			
; NUMBER OF SEQ ID NOS: 11			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 8			
; LENGTH: 525			
; TYPE: PRT			
; ORGANISM: Exophiala spinifera			
US-09-677-488A-8			
Query Match 100.0%; Score 2692; DB 20; Length 525;			
Best Local Similarity 100.0%; Pred. No. 3.8e-244;			
Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	MDSRPSGYGEGGTQTTKNTETAAGGASESLNVPLEKKQFGTTITVSLAFVICSWSAG	60
Db	1	MDSRPSGYGEGGTQTTKNTETAAGGASESLNVPLEKKQFGTTITVSLAFVICSWSAG	60
QY	61	ISGSLOALLAGGPTVLLYGLISTLVYICIAFSLAELTSVYPTAGQYHFASILAPKSI	120
Db	61	ISGSLOALLAGGPTVLLYGLISTLVYICIAFSLAELTSVYPTAGQYHFASILAPKSI	120
QY	121	NRSISYVCGVLSLWSIAIGSSVTMIPAQQIPALIAAASHYTSQDSWHVFLIYEGVALVY	180
Db	121	NRSISYVCGVLSLWSIAIGSSVTMIPAQQIPALIAAASHYTSQDSWHVFLIYEGVALVY	180
QY	181	LFNLFLKRNPNWVHEIGFGLTIALFVIFSIATLARSNPKAPNSQVWTWSFARDN	240
Db	181	LFNLFLKRNPNWVHEIGFGLTIALFVIFSIATLARSNPKAPNSQVWTWSFARDN	240
QY	241	CFILGLSTSCFMFICGLDAAMHIAEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGIT	300
Db	241	CFILGLSTSCFMFICGLDAAMHIAEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGIT	300
QY	301	DLDSILSSAGYIPFETMTQSLRSLSFATVLSGCGIVMAFFALNAVQETASRLTWSFARDN	360
Db	301	DLDSILSSAGYIPFETMTQSLRSLSFATVLSGCGIVMAFFALNAVQETASRLTWSFARDN	360
QY	361	GLVFSTHLERHPRWQVPVWSLFAFWGILATCGCIFLGSSTAFNALVNSAVVLOQLSFLI	420
Db	361	GLVFSTHLERHPRWQVPVWSLFAFWGILATCGCIFLGSSTAFNALVNSAVVLOQLSFLI	420
QY	421	PIALLLYOKRDPKFLPSTRAFLVPRGIGFLVNLAVVTSVTVFFSFPPLTVPTAASTMN	480
Db	421	PIALLLYOKRDPKFLPSTRAFLVPRGIGFLVNLAVVTSVTVFFSFPPLTVPTAASTMN	480
QY	481	YTSIIIGVALALGVNLVNVVHARKHYQGPHLELDGRVVGAEFQVGP	525
Db	481	YTSIIIGVALALGVNLVNVVHARKHYQGPHLELDGRVVGAEFQVGP	525
RESULT 3			
US-09-677-682A-8			
; Sequence 8, Application US/09677682A			
; GENERAL INFORMATION:			
; APPLICANT: Duvick, Jon			
; APPLICANT: Maddox, Joyce			
; APPLICANT: Gilliam, Jacob			
; APPLICANT: Folkerts, Otto			
; APPLICANT: Crasta, Oswald R.			
; TITLE OF INVENTION: Compositions and Methods for Fumonisin			
; TITLE OF INVENTION: Detoxification			
; FILE REFERENCE: 35718/204101			
; CURRENT APPLICATION NUMBER: US/09/677,682A			
; PRIOR FILING DATE: 2000-10-02			
; PRIOR FILING DATE: 1999-07-12			
; NUMBER OF SEQ ID NOS: 11			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 8			
; LENGTH: 525			
; TYPE: PRT			
; ORGANISM: Exophiala spinifera			
US-09-677-682A-8			
Query Match 100.0%; Score 2692; DB 20; Length 525;			
Best Local Similarity 100.0%; Pred. No. 3.8e-244;			
Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	MDSRPSGYGEGGTQTTKNTETAAGGASESLNVPLEKKQFGTTITVSLAFVICSWSAG	60
Db	1	MDSRPSGYGEGGTQTTKNTETAAGGASESLNVPLEKKQFGTTITVSLAFVICSWSAG	60
QY	61	ISGSLOALLAGGPTVLLYGLISTLVYICIAFSLAELTSVYPTAGQYHFASILAPKSI	120
Db	61	ISGSLOALLAGGPTVLLYGLISTLVYICIAFSLAELTSVYPTAGQYHFASILAPKSI	120
QY	121	NRSISYVCGVLSLWSIAIGSSVTMIPAQQIPALIAAASHYTSQDSWHVFLIYEGVALVY	180
Db	121	NRSISYVCGVLSLWSIAIGSSVTMIPAQQIPALIAAASHYTSQDSWHVFLIYEGVALVY	180
QY	181	LFNLFLKRNPNWVHEIGFGLTIALFVIFSIATLARSNPKAPNSQVWTWSFARDN	240
Db	181	LFNLFLKRNPNWVHEIGFGLTIALFVIFSIATLARSNPKAPNSQVWTWSFARDN	240
QY	241	CFILGLSTSCFMFICGLDAAMHIAEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGIT	300
Db	241	CFILGLSTSCFMFICGLDAAMHIAEECTDAARTVPKAVVSAIIIGFCTAFPTYIAVLYGIT	300

```
|||||
Db 241 CFILGLSTSCFMFIGLDRAMHAECDAAATVPKAVVSAILIGCFAPPTIAVLGIT 300
QY 301 DLDSILSSAGYIPFETMTQSLRSLSFATVLSGCGIVMAFFALNAVOETASRLTWSFARN 360
Db 301 DLDSILSSAGYIPFETMTQSLRSLSFATVLSGCGIVMAFFALNAVOETASRLTWSFARN 360
QY 361 GLVFSTHLERIHPRWQVPVWSLFATWGLILATCCGIFLGSSTAFNALVNSAVVLIQOLSFLI 420
Db 361 GLVFSTHLERIHPRWQVPVWSLFATWGLILATCCGIFLGSSTAFNALVNSAVVLIQOLSFLI 420
QY 421 PIALLLYQKRDPKFLPSTRAFLVPRGIGFLVNVAVVFTSVTTVFFSPFLTPVPTAASTMN 480
Db 421 PIALLLYQKRDPKFLPSTRAFLVPRGIGFLVNVAVVFTSVTTVFFSPFLTPVPTAASTMN 480
QY 481 YTSAILGVALALGVNLVNVVHARKHYQGPHLELDGRVVGAEFQVGP 525
Db 481 YTSAILGVALALGVNLVNVVHARKHYQGPHLELDGRVVGAEFQVGP 525
```

RESULT 4

```
US-09-677-682B-8
; Sequence 8, Application US/09677682B
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 35718/204101
; CURRENT APPLICATION NUMBER: US/09/677,682B
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-677-682B-8
```

```
Query Match 100.0%; Score 2692; DB 20; Length 525;
Best Local Similarity 100.0%; Pred. No. 3.8e-244;
Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDSRPSGYGEGKGTROTNTKTETAAAGGASESLNVPLEKKQFGTTIIVSLAFVICNSWAG 60
Db 1 MDSRPSGYGEGKGTROTNTKTETAAAGGASESLNVPLEKKQFGTTIIVSLAFVICNSWAG 60
QY 61 IGSGLQALLAGGPVTLVYICIAFSLAELTSSVPTAGGQYHFASILAPKSI 120
Db 61 IGSGLQALLAGGPVTLVYICIAFSLAELTSSVPTAGGQYHFASILAPKSI 120
QY 121 NRSISYVCGVLSLSWIAIGSSVTMIPAQQIPALIAAYSHYTSQDSWHVFLIYEGVALV 180
Db 121 NRSISYVCGVLSLSWIAIGSSVTMIPAQQIPALIAAYSHYTSQDSWHVFLIYEGVALV 180
QY 181 LFNLFALKRNPWVHEIGFGLTIALFVLSFTAILARSNPKAPNSQVMTAWSNYTGSVDG 240
Db 181 LFNLFALKRNPWVHEIGFGLTIALFVLSFTAILARSNPKAPNSQVMTAWSNYTGSVDG 240
QY 241 CFILGLSTSCFMFIGLDRAMHAECDAAATVPKAVVSAILIGCFAPPTIAVLGIT 300
Db 241 CFILGLSTSCFMFIGLDRAMHAECDAAATVPKAVVSAILIGCFAPPTIAVLGIT 300
QY 301 DLDSILSSAGYIPFETMTQSLRSLSFATVLSGCGIVMAFFALNAVOETASRLTWSFARN 360
Db 301 DLDSILSSAGYIPFETMTQSLRSLSFATVLSGCGIVMAFFALNAVOETASRLTWSFARN 360
```

```
QY 361 GLVFSTHLERIHPRWQVPVWSLFATWGLILATCCGIFLGSSTAFNALVNSAVVLIQOLSFLI 420
Db 361 GLVFSTHLERIHPRWQVPVWSLFATWGLILATCCGIFLGSSTAFNALVNSAVVLIQOLSFLI 420
QY 421 PIALLLYQKRDPKFLPSTRAFLVPRGIGFLVNVAVVFTSVTTVFFSPFLTPVPTAASTMN 480
Db 421 PIALLLYQKRDPKFLPSTRAFLVPRGIGFLVNVAVVFTSVTTVFFSPFLTPVPTAASTMN 480
QY 481 YTSAILGVALALGVNLVNVVHARKHYQGPHLELDGRVVGAEFQVGP 525
Db 481 YTSAILGVALALGVNLVNVVHARKHYQGPHLELDGRVVGAEFQVGP 525

RESULT 5
US-09-882-694-8
; Sequence 8, Application US/09882694
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 35718/208255
; CURRENT APPLICATION NUMBER: US/09/882,694
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-882-694-8
```

```
Query Match 100.0%; Score 2692; DB 22; Length 525;
Best Local Similarity 100.0%; Pred. No. 3.8e-244;
Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDSRPSGYGEGKGTROTNTKTETAAAGGASESLNVPLEKKQFGTTIIVSLAFVICNSWAG 60
Db 1 MDSRPSGYGEGKGTROTNTKTETAAAGGASESLNVPLEKKQFGTTIIVSLAFVICNSWAG 60
QY 61 IGSGLQALLAGGPVTLVYICIAFSLAELTSSVPTAGGQYHFASILAPKSI 120
Db 61 IGSGLQALLAGGPVTLVYICIAFSLAELTSSVPTAGGQYHFASILAPKSI 120
QY 121 NRSISYVCGVLSLSWIAIGSSVTMIPAQQIPALIAAYSHYTSQDSWHVFLIYEGVALV 180
Db 121 NRSISYVCGVLSLSWIAIGSSVTMIPAQQIPALIAAYSHYTSQDSWHVFLIYEGVALV 180
QY 181 LFNLFALKRNPWVHEIGFGLTIALFVLSFTAILARSNPKAPNSQVMTAWSNYTGSVDG 240
Db 181 LFNLFALKRNPWVHEIGFGLTIALFVLSFTAILARSNPKAPNSQVMTAWSNYTGSVDG 240
QY 241 CFILGLSTSCFMFIGLDRAMHAECDAAATVPKAVVSAILIGCFAPPTIAVLGIT 300
Db 241 CFILGLSTSCFMFIGLDRAMHAECDAAATVPKAVVSAILIGCFAPPTIAVLGIT 300
QY 301 DLDSILSSAGYIPFETMTQSLRSLSFATVLSGCGIVMAFFALNAVOETASRLTWSFARN 360
Db 301 DLDSILSSAGYIPFETMTQSLRSLSFATVLSGCGIVMAFFALNAVOETASRLTWSFARN 360
QY 361 GLVFSTHLERIHPRWQVPVWSLFATWGLILATCCGIFLGSSTAFNALVNSAVVLIQOLSFLI 420
Db 361 GLVFSTHLERIHPRWQVPVWSLFATWGLILATCCGIFLGSSTAFNALVNSAVVLIQOLSFLI 420
QY 421 PIALLLYQKRDPKFLPSTRAFLVPRGIGFLVNVAVVFTSVTTVFFSPFLTPVPTAASTMN 480
Db 421 PIALLLYQKRDPKFLPSTRAFLVPRGIGFLVNVAVVFTSVTTVFFSPFLTPVPTAASTMN 480
```

QY 481 YTSAGVALGVLNWNVHARKHYQGPHLELDGRVVGAEFOVGP 525
|||||
Db 481 YTSAGVALGVLNWNVHARKHYQGPHLELDGRVVGAEFOVGP 525

RESULT 6

US-09-882-694A-8
; Sequence 8, Application US/09882694A
; GENERAL INFORMATION:
; APPLICANT: Duwick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/882.694A
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 09/351,224
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-882-694A-8

Query Match 100.0%; Score 2692; DB 22; Length 525;
Best Local Similarity 100.0%; Pred. No. 3.8e-244;
Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDSRPSGYGEGKGTQTTKNTETAAAGGASESLNVPLEKKQFGTTIVSLAFVICSNWAG 60
|||||
Db 1 MDSRPSGYGEGKGTQTTKNTETAAAGGASESLNVPLEKKQFGTTIVSLAFVICSNWAG 60
QY 61 ISGSLQALLAGGPVTLTYGILSTLVYICIAFSLAELTSVYPTAGGYHFASILAPKSI 120
|||||
Db 61 ISGSLQALLAGGPVTLTYGILSTLVYICIAFSLAELTSVYPTAGGYHFASILAPKSI 120
QY 121 NRSISYVCGVLSLWSIAIGSSVTMIPAQQIPALIAAASHYTSQDSWHVFLIYEGVALV 180
|||||
Db 121 NRSISYVCGVLSLWSIAIGSSVTMIPAQQIPALIAAASHYTSQDSWHVFLIYEGVALV 180
QY 181 LFNLFALKRNPWVHEIGFGLTIALFVTSFTAILARSNPKAPNSQVWTAWSNYTGWSDGV 240
|||||
Db 181 LFNLFALKRNPWVHEIGFGLTIALFVTSFTAILARSNPKAPNSQVWTAWSNYTGWSDGV 240
QY 241 CFILGLSTCFMFICLDAAMHLAEECTDAARTVPKAVVSAIIIGFCTAPPYTTIAVLGIT 300
|||||
Db 241 CFILGLSTCFMFICLDAAMHLAEECTDAARTVPKAVVSAIIIGFCTAPPYTTIAVLGIT 300
QY 301 DLDSILSSAGYIPFETMTQSLRSLSFATVLSGGIVMAFFALNAVOETASRLTWSFARDN 360
|||||
Db 301 DLDSILSSAGYIPFETMTQSLRSLSFATVLSGGIVMAFFALNAVOETASRLTWSFARDN 360
QY 361 GLVFSTHLERIHPRMQVPVNSLFATWGLATCGCIFLGSSTAFNALVNSAVVLLQOLSFLI 420
|||||
Db 361 GLVFSTHLERIHPRMQVPVNSLFATWGLATCGCIFLGSSTAFNALVNSAVVLLQOLSFLI 420
QY 421 PIALLLYQKRDPKFLPSTRAFVLPGRGIFLVNVLAVVFTSVTTFFSFPLVPTAASTMN 480
|||||
Db 421 PIALLLYQKRDPKFLPSTRAFVLPGRGIFLVNVLAVVFTSVTTFFSFPLVPTAASTMN 480
QY 481 YTSAGVALGVLNWNVHARKHYQGPHLELDGRVVGAEFOVGP 525
|||||
Db 481 YTSAGVALGVLNWNVHARKHYQGPHLELDGRVVGAEFOVGP 525

RESULT 7

US-09-351-224-8
; Sequence 8, Application US/09351224
; GENERAL INFORMATION:
; APPLICANT: Duwick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fumonisin Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/351,224
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: permease
US-09-351-224-8

Query Match 98.6%; Score 2655; DB 17; Length 525;
Best Local Similarity 98.9%; Pred. No. 1.2e-240;
Matches 519; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 MDSRPSGYGEGKGTQTTKNTETAAAGGASESLNVPLEKKQFGTTIVSLAFVICSNWAG 60
|||||
Db 1 MDSRPSGYGEGKGTQTTKNTETAAAGGASESLNVPLEKKQFGTTIVSLAFVICSNWAG 60
QY 61 ISGSLQALLAGGPVTLTYGILSTLVYICIAFSLAELTSVYPTAGGYHFASILAPKSI 120
|||||
Db 61 ISGSLQALLAGGPVTLTYGILSTLVYICIAFSLAELTSVYPTAGGYHFASILAPKSI 120
QY 121 NRSISYVCGVLSLWSIAIGSSVTMIPAQQIPALIAAASHYTSQDSWHVFLIYEGVALV 180
|||||
Db 121 NRSISYVCGVLSLWSIAIGSSVTMIPAQQIPALIAAASHYTSQDSWHVFLIYEGVALV 180
QY 181 LFNLFALKRNPWVHEIGFGLTIALFVTSFTAILARSNPKAPNSQVWTAWSNYTGWSDGV 240
|||||
Db 181 LFNLFALKRNPWVHEIGFGLTIALFVTSFTAILARSNPKAPNSQVWTAWSNYTGWSDGV 240
QY 241 CFILGLSTCFMFICLDAAMHLAEECTDAARTVPKAVVSAIIIGFCTAPPYTTIAVLGIT 300
|||||
Db 241 CFILGLSTCFMFICLDAAMHLAEECTDAARTVPKAVVSAIIIGFCTAPPYTTIAVLGIT 300
QY 301 DLDSILSSAGYIPFETMTQSLRSLSFATVLSGGIVMAFFALNAVOETASRLTWSFARDN 360
|||||
Db 301 DLDSILSSAGYIPFETMTQSLRSLSFATVLSGGIVMAFFALNAVOETASRLTWSFARDN 360
QY 361 GLVFSTHLERIHPRMQVPVNSLFATWGLATCGCIFLGSSTAFNALVNSAVVLLQOLSFLI 420
|||||
Db 361 GLVFSTHLERIHPRMQVPVNSLFATWGLATCGCIFLGSSTAFNALVNSAVVLLQOLSFLI 420
QY 421 PIALLLYQKRDPKFLPSTRAFVLPGRGIFLVNVLAVVFTSVTTFFSFPLVPTAASTMN 480
|||||
Db 421 PIALLLYQKRDPKFLPSTRAFVLPGRGIFLVNVLAVVFTSVTTFFSFPLVPTAASTMN 480
QY 481 YTSAGVALGVLNWNVHARKHYQGPHLELDGRVVGAEFOVGP 525
|||||
Db 481 YTSAGVALGVLNWNVHARKHYQGPHLELDGRVVGAEFOVGP 525

RESULT 8

US-09-351-823-8
; Sequence 8, Application US/09351823
; GENERAL INFORMATION:
; APPLICANT: Duwick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fumonisin Detoxification
; FILE REFERENCE: 5718-56
; CURRENT APPLICATION NUMBER: US/09/351,823
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 525

; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: permease
US-09-351-823-8

Query Match 98.6%; Score 2655; DB 17; Length 525;
Best Local Similarity 98.9%; Pred. No. 1.2e-240;
Matches 519; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 MDSRPSGYGKGGTROTTKNTETAAGGASESLNVPLEKKQFGCTITIVSLAFVICSNWAG 60
DB 1 MDSRPSGYGLKGGTROTTKNTETAAGGASESLNVPLEKKQFGCTITIVSLAFVICSNWAG 60
QY 61 ISGSLQALLAGGPPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGQYHFASILAPKSI 120
DB 61 ISGSLQALLAGGPPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGQYHFASILAPKSI 120
QY 121 NRSISYVCGLVSLLSWIAIGSSVTMIPAQQIPALIAAAYSHYTSQDSWHVFLIYEGVALV 180
DB 121 NRSISYVCGLVSLLSWIAIGSSVTMIPAQQIPALIAAAYSHYTSQDSWHVFLIYEGVALV 180
QY 181 LFNLFALKRNPWVHEIGFGLTIALFVISTAILARSNPKAPNSQVMTAWSNYTGSVDGV 240
DB 181 LFNLFALKRNPWVHEIGFGLTIALFVISTAILARSNPKAPNSQVMTAWSNYTGSVDGV 240
QY 241 CFILGLSTSCFMFIGLDAAMHLAEECTDAARTVPKAVVSAILIIGCTAPPTIAYLYGIT 300
DB 241 CFILGLSTSCFMFIGLDAAMHLAEECTDAARTVPKAVVSAILIIGCTAPPTIAYLYGIT 300
QY 301 DLDSILSSAGYIPFETMQRSARIRLFATVLSGGGIVMAFFALNAVQETASRLTWSFARDN 360
DB 301 DLDSILSSAGYIPFETMQRSARIRLFATVLSGGGIVMAFFALNAVQETASRLTWSFARDN 360
QY 361 GLVFSTHLERIHPRMOPVMSLFAFWGILATCGCIFLGSSTAFNALVNSAVVLOQLSFLI 420
DB 361 GLVFSTHLERIHPRMOPVMSLFAFWGILATCGCIFLGSSTAFNALVNSAVVLOQLSFLI 420
QY 421 PIALLYQKRPKFLPSTRAFLVPRGIGFLVNLAVVTSVTVFFSPFLVPTAASTMN 480
DB 421 PIALLYQKRPKFLPSTRAFLVPRGIGFLVNLAVVTSVTVFFSPFLVPTAASTMN 480
QY 481 YTSAILGVALALGVNLVNVVHARKHYQGPHELDGRVVGAEFOVGP 525
DB 481 YTSAILGVALALGVNLVNVVHARKHYQGPHELDGRVVGAEFOVGP 525

RESULT 9
US-09-677-488-8
; Sequence 8, Application US/09677488
; GENERAL INFORMATION:
; APPLICANT: Duwick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fummonisin Detoxification
; FILE REFERENCE: 35718/204100 (5718-111A)
; CURRENT APPLICATION NUMBER: US/09/677,488
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: permease, translation of fully spliced cDNA
US-09-677-488-8

Query Match 98.6%; Score 2655; DB 20; Length 525;
Best Local Similarity 98.9%; Pred. No. 1.2e-240;
Matches 519; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 MDSRPSGYGKGGTROTTKNTETAAGGASESLNVPLEKKQFGCTITIVSLAFVICSNWAG 60
DB 1 MDSRPSGYGLKGGTROTTKNTETAAGGASESLNVPLEKKQFGCTITIVSLAFVICSNWAG 60
QY 61 ISGSLQALLAGGPPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGQYHFASILAPKSI 120
DB 61 ISGSLQALLAGGPPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGQYHFASILAPKSI 120
QY 121 NRSISYVCGLVSLLSWIAIGSSVTMIPAQQIPALIAAAYSHYTSQDSWHVFLIYEGVALV 180
DB 121 NRSISYVCGLVSLLSWIAIGSSVTMIPAQQIPALIAAAYSHYTSQDSWHVFLIYEGVALV 180
QY 181 LFNLFALKRNPWVHEIGFGLTIALFVISTAILARSNPKAPNSQVMTAWSNYTGSVDGV 240
DB 181 LFNLFALKRNPWVHEIGFGLTIALFVISTAILARSNPKAPNSQVMTAWSNYTGSVDGV 240
QY 241 CFILGLSTSCFMFIGLDAAMHLAEECTDAARTVPKAVVSAILIIGCTAPPTIAYLYGIT 300
DB 241 CFILGLSTSCFMFIGLDAAMHLAEECTDAARTVPKAVVSAILIIGCTAPPTIAYLYGIT 300
QY 301 DLDSILSSAGYIPFETMQRSARIRLFATVLSGGGIVMAFFALNAVQETASRLTWSFARDN 360
DB 301 DLDSILSSAGYIPFETMQRSARIRLFATVLSGGGIVMAFFALNAVQETASRLTWSFARDN 360
QY 361 GLVFSTHLERIHPRMOPVMSLFAFWGILATCGCIFLGSSTAFNALVNSAVVLOQLSFLI 420
DB 361 GLVFSTHLERIHPRMOPVMSLFAFWGILATCGCIFLGSSTAFNALVNSAVVLOQLSFLI 420
QY 421 PIALLYQKRPKFLPSTRAFLVPRGIGFLVNLAVVTSVTVFFSPFLVPTAASTMN 480
DB 421 PIALLYQKRPKFLPSTRAFLVPRGIGFLVNLAVVTSVTVFFSPFLVPTAASTMN 480
QY 481 YTSAILGVALALGVNLVNVVHARKHYQGPHELDGRVVGAEFOVGP 525
DB 481 YTSAILGVALALGVNLVNVVHARKHYQGPHELDGRVVGAEFOVGP 525

RESULT 10
US-09-677-682-8
; Sequence 8, Application US/09677682
; GENERAL INFORMATION:
; APPLICANT: Duwick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fummonisin Detoxification
; FILE REFERENCE: 35718/204101 (5718-111B)
; CURRENT APPLICATION NUMBER: US/09/677,682
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: permease, translation of fully spliced cDNA
US-09-677-682-8

Query Match 98.6%; Score 2655; DB 20; Length 525;
Best Local Similarity 98.9%; Pred. No. 1.2e-240;
Matches 519; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 MDSRPSGYGKGGTROTTKNTETAAGGASESLNVPLEKKQFGCTITIVSLAFVICSNWAG 60
DB 1 MDSRPSGYGLKGGTROTTKNTETAAGGASESLNVPLEKKQFGCTITIVSLAFVICSNWAG 60
QY 61 ISGSLQALLAGGPPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGQYHFASILAPKSI 120
DB 61 ISGSLQALLAGGPPVTLTYGILISTLVYICIAFSLAELTSVYPTAGGQYHFASILAPKSI 120
QY 121 NRSISYVCGLVSLLSWIAIGSSVTMIPAQQIPALIAAAYSHYTSQDSWHVFLIYEGVALV 180

Db 121 NRSISVYCGVLSLWSIAIGSSVTMIPAQQIPALIAAASHYTSQDSWHVFLIYEGVALV 180
QY 181 LLENFLALKRNPVWHEIGFGTLTALFVISTAILARSNKAPNSQVWTAHNSYTGSDGV 240
Db 181 LLENFLALKRNPVWHEIGFGTLTALFVISTAILARSNKAPNSQVWTAHNSYTGSDGV 240
QY 241 CFILGLSTCFMFIGLDAAMHLAEECTDAARTVPKAVVSIIIGFCTAFPTTAVLYGIT 300
Db 241 CFILGLSTCFMFIGLDAAMHLAEECTDAARTVPKAVVSIIIGFCTAFPTTAVLYGIT 300
QY 301 DLDSILSSAGYIPFTMTQSLRSLSFATVLSGCGIVMAFFALNAVQETASRLTWSFARDN 360
Db 301 DLDSILSSAGYIPFTMTQSLRSLSFATVLSGCGIVMAFFALNAVQETASRLTWSFARDN 360
QY 361 GLVFSFHLERIPRQVVPVWSLFAATWCILATCGCIFLGSSTAFNALVNSAVVLOQLSFLI 420
Db 361 GLVFSFHLERIPRQVVPVWSLFAATWCILATCGCIFLGSSTAFNALVNSAVVLOQLSFLI 420
QY 421 PTALLLYQKRDPKFLPSTRAFLVPRGIGFLVNVLAUVFTSVTVFFSFPLTVPTAASTMN 480
Db 421 PTALLLYQKRDPKFLPSTRAFLVPRGIGFLVNVLAUVFTSVTVFFSFPLTVPTAASTMN 480
QY 481 YTSAILGVALGVNLVWVHARKHYOGPHLELDGRVVGAEFQVGP 525
Db 481 YTSAILGVALGVNLVWVHARKHYOGPHLELDGRVVGAEFQVGP 525

RESULT 11
US-09-248-796-20096
; Sequence 20096, Application US/09248796
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796
; CURRENT FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 20096
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796-20096

Query Match 20.9%; Score 561.5; DB 16; Length 585;
Best Local Similarity 29.9%; Pred. No. 2e-43;
Matches 152; Conservative 106; Mismatches 210; Indels 41; Gaps 13;
QY 17 TTKNTETAAAGGASESLNVPLEKKQFGTITIVSLAFVICSNWAGISGSLQALLAGGPVT 76
Db 88 TSKDEDTKMS-----RNFSLTQIVTIGFGLTNSMLGILSTLILSISSLGPLL 134
QY 77 LLYGILISTLVYICIAFSLAELTSVYTAGQYHFASTILAPKINSRISYVCGLVLSLWS 136
Db 135 VVYGLIVASVCIATVLTGEMALAMPAGGQYVWARVAPKYSFLAYITGSIWSGGA 194
QY 137 IATGSSVTMIPAQQIPALIAASHYTSQDSWHVFLIYEGVALVNLFLALKRNPVWHE 196
Db 195 IFTTASMLAVAYQVLFWMNTHPDHVNQKEVFIYINLWILFFFNH-RFLPMIGD 253
QY 197 IGFGTLIALFVISTAILARSNKAPNSQ-VWTAMSNYTGW-SDGVCFTILGSTSCPMFI 254
Db 254 SVFGISLTSYCIILITVLVCARGHYQDAKFVHVHANNVTGWPSKGIAPTVGLVNPWAFS 313
QY 255 GLDAAMHLAEECTDAARTVPKAVVSIIIGFCTAFPTTAVLYGITDLDLSLSSA-GYIP 313
Db 314 CLDSVTHLSEETAHPERDVPRAVLSTVGIGFTVSTFSIAMFFCIRNLEEINMSATGPPM 373
QY 314 FETMTQSLRSLSFATVLSGCGIVMAFFALNAVQETAS-----RLTWSFARDNGLVFSTH 367
Db 374 LDIFYQALGSTKVGAII--CLGSLITLVATGC---TLSCIIYQLRLLFSFLRDNMPLSKY 428
QY 368 LERHPRQVVPVWS-LFATWGLILATCGCIFLGSSTAFNALVNSAVVLOQLSFLIPIALL 426
Db 429 LSIIDPKTGGPGFYAHLFST--VIVSVSVLVFSDAALQATALACVSFLLIAYLPTICLL 486
QY 427 YOKRDPKFLPSTRAFLVPRGIGFLVNVLAUVFTSVTVFFSFPLTVPTAASTMNTSII 486
Db 487 ARRRQIRHGP-----FWLQK-IGVFCNFVLLAWCIFAUVFFSFPPANYPTAEGMNYFCVVL 541
QY 487 GVALALGVNLVWVHARKH-----YQGPHL 510
Db 542 VVYIICMLGYWFWPIKKYACKYNFRGNNL 570

QY 368 LERHPRQVVPVWS-LFATWGLILATCGCIFLGSSTAFNALVNSAVVLOQLSFLIPIALL 426
Db 429 LSIIDPKTGGPGFYAHLFST--VIVSVSVLVFSDAALQATALACVSFLLIAYLPTICLL 486
QY 427 YOKRDPKFLPSTRAFLVPRGIGFLVNVLAUVFTSVTVFFSFPLTVPTAASTMNTSII 486
Db 487 ARRRQIRHGP-----FWLQK-IGVFCNFVLLAWCIFAUVFFSFPPANYPTAEGMNYFCVVL 541
QY 487 GVALALGVNLVWVHARKH-----YQGPHL 510
Db 542 VVYIICMLGYWFWPIKKYACKYNFRGNNL 570
RESULT 12
US-60-096-409-20096
; Sequence 20096, Application US/60096409A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALB
; FILE REFERENCE: GTC-016P
; CURRENT APPLICATION NUMBER: US/60/096.409A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 20096
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Candida albicans
US-60-096-409-20096

Query Match 20.9%; Score 561.5; DB 26; Length 585;
Best Local Similarity 29.9%; Pred. No. 2e-43;
Matches 152; Conservative 106; Mismatches 210; Indels 41; Gaps 13;
QY 17 TTKNTETAAAGGASESLNVPLEKKQFGTITIVSLAFVICSNWAGISGSLQALLAGGPVT 76
Db 88 TSKDEDTKMS-----RNFSLTQIVTIGFGLTNSMLGILSTLILSISSLGPLL 134
QY 77 LLYGILISTLVYICIAFSLAELTSVYTAGQYHFASTILAPKINSRISYVCGLVLSLWS 136
Db 135 VVYGLIVASVCIATVLTGEMALAMPAGGQYVWARVAPKYSFLAYITGSIWSGGA 194
QY 137 IATGSSVTMIPAQQIPALIAASHYTSQDSWHVFLIYEGVALVNLFLALKRNPVWHE 196
Db 195 IFTTASMLAVAYQVLFWMNTHPDHVNQKEVFIYINLWILFFFNH-RFLPMIGD 253
QY 197 IGFGTLIALFVISTAILARSNKAPNSQ-VWTAMSNYTGW-SDGVCFTILGSTSCPMFI 254
Db 254 SVFGISLTSYCIILITVLVCARGHYQDAKFVHVHANNVTGWPSKGIAPTVGLVNPWAFS 313
QY 255 GLDAAMHLAEECTDAARTVPKAVVSIIIGFCTAFPTTAVLYGITDLDLSLSSA-GYIP 313
Db 314 CLDSVTHLSEETAHPERDVPRAVLSTVGIGFTVSTFSIAMFFCIRNLEEINMSATGPPM 373
QY 314 FETMTQSLRSLSFATVLSGCGIVMAFFALNAVQETAS-----RLTWSFARDNGLVFSTH 367
Db 374 LDIFYQALGSTKVGAII--CLGSLITLVATGC---TLSCIIYQLRLLFSFLRDNMPLSKY 428
QY 368 LERHPRQVVPVWS-LFATWGLILATCGCIFLGSSTAFNALVNSAVVLOQLSFLIPIALL 426
Db 429 LSIIDPKTGGPGFYAHLFST--VIVSVSVLVFSDAALQATALACVSFLLIAYLPTICLL 486
QY 427 YOKRDPKFLPSTRAFLVPRGIGFLVNVLAUVFTSVTVFFSFPLTVPTAASTMNTSII 486
Db 487 ARRRQIRHGP-----FWLQK-IGVFCNFVLLAWCIFAUVFFSFPPANYPTAEGMNYFCVVL 541
QY 487 GVALALGVNLVWVHARKH-----YQGPHL 510
Db 542 VVYIICMLGYWFWPIKKYACKYNFRGNNL 570

```

RESULT 13
PCT-US02-03987-15766
; Sequence 15766, Application PC/TUS0203987
; GENERAL INFORMATION:
; APPLICANT: Elittra Pharmaceuticals, Inc.
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibits C
; FILE REFERENCE: ELITRA.028VPC
; CURRENT APPLICATION NUMBER: PCT/US02/03987
; CURRENT FILING DATE: 2002-02-02
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 15811
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15766
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Candida albicans
PCT-US02-03987-15766

Query Match      20.5%; Score 552; DB 1; Length 503;
Best Local Similarity 29.7%; Pred. No. 1.3e-42;
Matches 143; Conservative 96; Mismatches 215; Indels 28; Gaps 11;

QY 39 KKQFGTITIVSLAFVICSNWAGISGSLQALLAGGPVTLVYLGILISTLVYICIAFSLAEL 98
Db 26 EKNFLLSTCAQFTFLICSAIAIGTFTLSTVGVGSPVLIFGFIVATFDLIICYSLAEL 85
QY 99 TSVYTAGQVHFASILAPKSNRSISYVCGVLVSLLSWIAIGSSVTMIPAAQIPALIAAY 158
Db 86 ASAYPHSSAQIHWTVCLASEKYKRSLSFLTGILSCAGWIFACFSTYVASMFIILAAQIY 145
QY 159 SHYTSQDSWHVFLIYEGVALVLLFNLFALKRNPWVHEIGFGLTIALFVISP-----IA 212
Db 146 HQDYVPKSFHYLYLVAVFLSGYLVNVLKLLPLI-----TNISVAVINFGTFFIIT 199
QY 213 ILARSNKAPNSQVMTANSNTGW--SDGVCFTLIGLSTSCFMFIGLDAAMHIAEECTDAAR 271
Db 200 LLVKS-PKQSAEFVKNIINETGSSNGVVFGLMPLSLACVTFLFDGAVHLTDEIAQPER 258
QY 272 TVPKAVSAIIGCTAFPTIYAVLGITDLSLSSAGYIPF-ETWTQSLRSLSFATV- 329
Db 259 NIPLVMVINSNTLSGVMAFFAAIVMFCVVNVNLSNPVGGSEPIVQLMYDQSEALTIG 318
QY 330 LSCGGIVMAFFALNAVOETA-SRLTWSFARDNGLVFSTHLERIHPRQVPMVSLFATWGI 388
Db 319 VVC--LIITFVGSSVMYTTSLRLWSPANSNGLPFSKYIGEVSNNLSPVYALSFLTIVL 376
QY 389 LATCCGIFLGSSTAFNALVNSAVVLOOLSFILPIALTYQKRDPKFLPSTR-----AFVL 443
Db 377 CIIIGTLIMSGDGNALNVLGTSVMVCINLSYLIPIACLLVKS---KFSTTHRFNERPYFCL 433
QY 444 PRGIGFLVNLAVVFTSVTTFFSFPLTPVTAASMTNYSIAIGVALGALVNNVWVHARK 503
Db 434 GK-FGLPMNIASVLWVCFIMVWLNEPLSPYPTSDNMNYACVVLGTCIIGIILWFVHGRS 492
QY 504 HY 505
Db 493 RY 494

RESULT 14
US-10-032-585-7822
; Sequence 7822, Application US/10032585
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585

```

```

; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7822
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7822

Query Match      20.5%; Score 552; DB 24; Length 503;
Best Local Similarity 29.7%; Pred. No. 1.3e-42;
Matches 143; Conservative 96; Mismatches 215; Indels 28; Gaps 11;

QY 39 KKQFGTITIVSLAFVICSNWAGISGSLQALLAGGPVTLVYLGILISTLVYICIAFSLAEL 98
Db 26 EKNFLLSTCAQFTFLICSAIAIGTFTLSTVGVGSPVLIFGFIVATFDLIICYSLAEL 85
QY 99 TSVYTAGQVHFASILAPKSNRSISYVCGVLVSLLSWIAIGSSVTMIPAAQIPALIAAY 158
Db 86 ASAYPHSSAQIHWTVCLASEKYKRSLSFLTGILSCAGWIFACFSTYVASMFIILAAQIY 145
QY 159 SHYTSQDSWHVFLIYEGVALVLLFNLFALKRNPWVHEIGFGLTIALFVISP-----IA 212
Db 146 HQDYVPKSFHYLYLVAVFLSGYLVNVLKLLPLI-----TNISVAVINFGTFFIIT 199
QY 213 ILARSNKAPNSQVMTANSNTGW--SDGVCFTLIGLSTSCFMFIGLDAAMHIAEECTDAAR 271
Db 200 LLVKS-PKQSAEFVKNIINETGSSNGVVFGLMPLSLACVTFLFDGAVHLTDEIAQPER 258
QY 272 TVPKAVSAIIGCTAFPTIYAVLGITDLSLSSAGYIPF-ETWTQSLRSLSFATV- 329
Db 259 NIPLVMVINSNTLSGVMAFFAAIVMFCVVNVNLSNPVGGSEPIVQLMYDQSEALTIG 318
QY 330 LSCGGIVMAFFALNAVOETA-SRLTWSFARDNGLVFSTHLERIHPRQVPMVSLFATWGI 388
Db 319 VVC--LIITFVGSSVMYTTSLRLWSPANSNGLPFSKYIGEVSNNLSPVYALSFLTIVL 376
QY 389 LATCCGIFLGSSTAFNALVNSAVVLOOLSFILPIALTYQKRDPKFLPSTR-----AFVL 443
Db 377 CIIIGTLIMSGDGNALNVLGTSVMVCINLSYLIPIACLLVKS---KFSTTHRFNERPYFCL 433
QY 444 PRGIGFLVNLAVVFTSVTTFFSFPLTPVTAASMTNYSIAIGVALGALVNNVWVHARK 503
Db 434 GK-FGLPMNIASVLWVCFIMVWLNEPLSPYPTSDNMNYACVVLGTCIIGIILWFVHGRS 492
QY 504 HY 505
Db 493 RY 494

RESULT 15
US-10-072-851-15766
; Sequence 15766, Application US/10072851
; GENERAL INFORMATION:
; APPLICANT: Carr, Grant J.
; APPLICANT: Xu, H. Howard
; APPLICANT: Foulkes, J. Gordon
; APPLICANT: Zamudio, Carlos
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Roemer, Terry
; APPLICANT: Jiang, Bo
; APPLICANT: Boone, Charles
; APPLICANT: Bussey, Howard
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibit
; FILE REFERENCE: ELITRA.028A
; CURRENT APPLICATION NUMBER: US/10/072,851

```

; CURRENT FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 15811
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15766
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-072-851-15766

Query Match 20.5%; Score 552; DB 24; Length 503;
Best Local Similarity 29.7%; Pred. No. 1.3e-42;
Matches 143; Conservative 96; Mismatches 215; Indels 28; Gaps 11:

QY	39	KKQFGTITIVSLAFVICSNWAGISGLQALLAGGPPVTLTYGILISTVLVYICIAFSLAEL	98
DB	26	EKNFSLSTCAFOFTLICSALAIGTFLSTVIGVGSPVLIFGPIVATFDLIICYSLAEL	85
QY	99	TSVPTAGGQYHPASTILAPKSIHRSISYVCGIVSLLSWTAIGSSVTMPAQOIPALIAAY	158
DB	86	ASAPHSSAQIHWTYCLASEKYKRSLSFLTGILSCAGWIFACFSSTYVASMFLALAQIY	145
QY	159	SHYQSQSHVFLIYEGVALVLLFNLFALKRNPWVHEIGFGLTIALFVLSF-----IA	212
DB	146	HQDYVPKSFHYLVYLVAVFLSGYLVNVFLVKLLPLI-----TNISVAVINFGTFFIIT	199
QY	213	ILARSPKAPNSQVWTAWSYTGW-SDGVCFILGLSTSCPMFGLDAAMHAEECTDAAR	271
DB	200	LLVKS-PKQSAEFVKNIINETGSSNGVVFVFLGMLPSLACVTLFDCAVHLTDEIAQPER	258
QY	272	TPKAVVSAIIGFCTAFPTYIAVLXGTDLSILSAGVIPF-ETMTQSLRSLSFATV-	329
DB	259	NIPLVWVISTLSGVMAFFAAIYVMFCVNVNNSNPVCGEPIVQLMYDSFQSEALTIG	318
QY	330	LSCGGIVMAEFALNAVOETA-SRLTWSFARDNGLVFSTHLERIHPRWQVVPVWSLFATWGI	388
DB	319	VVC--LILTFVGSSYMYTSTRLWSFANSNGLPFSKYIGEVSNNLSPVYALSFLTVL	376
QY	389	LATCGCIFLGSSTAFNALVNSAVVLOOLSELIPALILLYOKRDPKPLPSTR-----AFVL	443
DB	377	CIITGLIMSGDGNALNAVLTGSMVCINLSYLIPIACLLVKS---KFSTTHRFNERPYECL	433
QY	444	PRGIGFLVNVYLVVFTSVTTVFESFPLTPTAASTNMYTSAIIGVALALGVLNWNVHARK	503
DB	434	GK-FGLPMNIASVLWVCFIMVWLNFLPSYPTSDNNMYACVVLGITCIIGIILWFVHGRS	492
QY	504	HY 505	
DB	493	RY 494	

Search completed: April 27, 2002, 07:55:22
Job time: 583 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:46:46 ; Search time 43.81 Seconds
(without alignments)
292.706 Million cell updates/sec

Title: US-09-882-694A-8

Perfect score: 2892
Sequence: 1 MDSRPSGVGKGTROTTKN.....QGPHELDGRVGAFOVGP 525

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/2/1aa/5A-COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B-COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A-COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B-COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS-COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	164	6.1	622	2	US-08-132-990A-4
2	164	6.1	622	5	PCT-US92-09382-4
3	145	5.4	629	2	US-08-132-990A-8
4	145	5.4	629	5	PCT-US92-09382-8
5	137.5	5.1	511	2	US-09-073-362-1
6	137.5	5.1	511	2	US-09-243-920-1
7	129.5	4.8	515	2	US-09-073-362-3
8	129.5	4.8	515	2	US-09-243-920-3
9	126.5	4.7	503	4	US-09-068-195-24
10	116.5	4.3	453	2	US-08-132-990A-6
11	116.5	4.3	453	5	PCT-US92-09382-6
12	114	4.2	493	1	US-08-362-512A-4
13	114	4.2	493	4	US-08-964-939-4
14	109.5	4.1	513	4	US-09-097-889-15
15	107	4.0	438	2	US-08-677-049-9
16	107	4.0	506	4	US-09-066-047-7
17	104	3.9	367	2	US-08-132-990A-2
18	104	3.9	367	5	PCT-US92-09382-2
19	104	3.9	485	1	US-08-362-512A-2
20	104	3.9	485	4	US-08-964-939-2
21	98.5	3.7	418	4	US-09-030-267-5
22	97.5	3.6	459	4	US-09-097-889-22
23	97	3.6	434	2	US-08-677-049-7
24	97	3.6	472	4	US-09-354-129-2
25	97	3.6	1212	4	US-09-268-866-2
26	96.5	3.6	429	2	US-08-677-049-5
27	95	3.5	635	2	US-09-014-969-11

28 94.5 3.5 1353 4 US-09-398-193-99 Sequence 99, Appli
29 94 3.5 1165 1 US-08-240-357-2 Sequence 2, Appli
30 94 3.5 1294 2 US-08-819-288-3 Sequence 3, Appli
31 94 3.5 1294 4 US-09-400-348-3 Sequence 3, Appli
32 94 3.5 1321 1 US-08-261-822A-3 Sequence 3, Appli
33 94 3.5 1321 5 PCT-US95-0774A-3 Sequence 13, Appli
34 93.5 3.5 584 2 US-08-928-692-13 Sequence 2, Appli
35 93 3.5 1443 1 US-08-108-872B-2 Sequence 2, Appli
36 92.5 3.4 416 4 US-09-333-208-2 Sequence 2, Appli
37 92.5 3.4 416 4 US-08-333-254-2 Sequence 3, Appli
38 92.5 3.4 1305 4 US-08-864-785-3 Sequence 2, Appli
39 92.5 3.4 1353 3 US-08-894-173-2 Sequence 2, Appli
40 92.5 3.4 1353 4 US-09-398-193-2 Sequence 2, Appli
41 90.5 3.4 294 1 US-08-142-439A-4 Sequence 4, Appli
42 90.5 3.4 294 2 US-08-869-477-4 Sequence 4, Appli
43 90.5 3.4 350 1 US-08-118-270-41 Sequence 41, Appli
44 90.5 3.4 350 5 PCT-US93-08528-41 Sequence 41, Appli
45 90.5 3.4 1454 4 US-08-392-459-22 Sequence 22, Appli

ALIGNMENTS

RESULT 1
US-08-132-990A-4
; Sequence 4, Application US/08132990A
; Patent No. 5834589
; GENERAL INFORMATION:
; APPLICANT: MERUELO, DANIEL
; APPLICANT: YOSHIMOTO, TAKAYUKI
; TITLE OF INVENTION: Human Retrovirus Receptor and DNA Coding Therefor
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/132,990A
; FILING DATE: 07-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/084,729
; FILING DATE: 29-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/05569
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/899,075
; FILING DATE: 11-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/806,178
; FILING DATE: 13-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/627,950
; FILING DATE: 14-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 8105-004-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66441 PENNIE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 622 amino acids

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-132-990A-4

Query Match 6.1%; Score 164; DB 2; Length 622;
Best Local Similarity 20.8%; Pred. No. 5.1e-08;
Matches 106; Conservative 84; Mismatches 202; Indels 118; Gaps 23;

```
QY 69 LLAG-----GPTLLYGLILSTLVYICIAFSLAELTSVYPTAGGQYHFASILAPKSN 121
Db 52 VLAGAVARENAGP-AIVISFLIAALASVLAGCYGEGFARVPTGSAIYLYSV---TVG 106
QY 122 RSISYVGLVSLLSWIAIGSSVTMIPAAQIPALIAAYSHYTSQDSWHVFLIYEG----- 175
Db 107 ELWAFITGNWLLSYIIGTSVARAWSATFDELICKPIGEFSRQ--HMALNAPGVLAQTP 164
QY 176 ----VALVVLLENFAL--KRNPNWHEIGFGLTIALFVIFSIAT-----LARS 217
Db 165 DIFAVIIIIITLGLTLGVKESAMVNI--FTCINVLVLCFIVVSGFVKSGIKNWLTEK 222
QY 218 NKPAPNSQVWTANSYNTGWSGVCFILGLSTSCF-MFIGLDAAMHLAEECTDAARTVPKA 276
Db 223 NFGCNNDNTNVKYGEGFMPFGSGVLGSAATCFYAFVGFDCIATTEGVEKNPKQAI 282
QY 277 VSAIIIGFCTAF-----PYTIAVLYGITDLSILSSAGYIPFETMTQSLRSLSFATVLS 332
Db 283 IVASLLICFIAYFGVSAALTLMPYFCLDIDSP-----PGAFKHOGWEAKYAVAI-- 334
QY 333 GGIVMAFFALNAVQETASRLTWSFARDNGLVFSTHLERIHPRMQVPVMSLFWTGLATC 392
Db 335 GSICALSTSLGSMFMPRVIYAMAED-GLLFK-FLAKINNRTKTPVIATVTSGAIAVM 392
QY 393 GCIF-----LGSSTAFNALVNSAVL-----QQLSFLPIA-----LL 425
Db 393 AFLFELKDLVDMISGTLTLLAYSLVAACVLVRYQEPQPNLVYQMARTTEELDRVDONELV 452
QY 426 LYQKRDPKFLP-----STRAFLVPRGI-----GFLNV-----LAVVFTSVTTVFFSFPL 470
Db 453 SASESQTGFLPVAEKFSLSKLSILSPKNVPSKFSGLIVNISAGLLAALIITVCIV----- 506
QY 471 TVPTAASMTNYSATIG-VALALGVLNVVV 499
Db 507 -----AVLGREALAEGLT-WAV 522
```

RESULT 2

PCT-US92-09382-4
; Sequence 4, Application PC/TUS9209382
; GENERAL INFORMATION:
; APPLICANT: MERUELO, DANIEL
; APPLICANT: YOSHIMOTO, TAKAYUKI
; TITLE OF INVENTION: Human Retrovirus Receptor and DNA Coding
; TITLE OF INVENTION: Therefor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/09382
; FILING DATE: 19921213
; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:
; NAME: Livnat, Shmuel
; REGISTRATION NUMBER: 33,949
; REFERENCE/DOCKET NUMBER: MERUELO-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 628-5197
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 622 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US92-09382-4

Query Match 6.1%; Score 164; DB 5; Length 622;
Best Local Similarity 20.8%; Pred. No. 5.1e-08;
Matches 106; Conservative 84; Mismatches 202; Indels 118; Gaps 23;

```
QY 69 LLAG-----GPTLLYGLILSTLVYICIAFSLAELTSVYPTAGGQYHFASILAPKSN 121
Db 52 VLAGAVARENAGP-AIVISFLIAALASVLAGCYGEGFARVPTGSAIYLYSV---TVG 106
QY 122 RSISYVGLVSLLSWIAIGSSVTMIPAAQIPALIAAYSHYTSQDSWHVFLIYEG----- 175
Db 107 ELWAFITGNWLLSYIIGTSVARAWSATFDELICKPIGEFSRQ--HMALNAPGVLAQTP 164
QY 176 ----VALVVLLENFAL--KRNPNWHEIGFGLTIALFVIFSIAT-----LARS 217
Db 165 DIFAVIIIIITLGLTLGVKESAMVNI--FTCINVLVLCFIVVSGFVKSGIKNWLTEK 222
QY 218 NKPAPNSQVWTANSYNTGWSGVCFILGLSTSCF-MFIGLDAAMHLAEECTDAARTVPKA 276
Db 223 NFGCNNDNTNVKYGEGFMPFGSGVLGSAATCFYAFVGFDCIATTEGVEKNPKQAI 282
QY 277 VSAIIIGFCTAF-----PYTIAVLYGITDLSILSSAGYIPFETMTQSLRSLSFATVLS 332
Db 283 IVASLLICFIAYFGVSAALTLMPYFCLDIDSP-----PGAFKHOGWEAKYAVAI-- 334
QY 333 GGIVMAFFALNAVQETASRLTWSFARDNGLVFSTHLERIHPRMQVPVMSLFWTGLATC 392
Db 335 GSICALSTSLGSMFMPRVIYAMAED-GLLFK-FLAKINNRTKTPVIATVTSGAIAVM 392
QY 393 GCIF-----LGSSTAFNALVNSAVL-----QQLSFLPIA-----LL 425
Db 393 AFLFELKDLVDMISGTLTLLAYSLVAACVLVRYQEPQPNLVYQMARTTEELDRVDONELV 452
QY 426 LYQKRDPKFLP-----STRAFLVPRGI-----GFLNV-----LAVVFTSVTTVFFSFPL 470
Db 453 SASESQTGFLPVAEKFSLSKLSILSPKNVPSKFSGLIVNISAGLLAALIITVCIV----- 506
QY 471 TVPTAASMTNYSATIG-VALALGVLNVVV 499
Db 507 -----AVLGREALAEGLT-WAV 522
```

RESULT 3

US-08-132-990A-8
; Sequence 8, Application US/08132990A
; Patent No. 5834589
; GENERAL INFORMATION:
; APPLICANT: MERUELO, DANIEL
; APPLICANT: YOSHIMOTO, TAKAYUKI
; TITLE OF INVENTION: Human Retrovirus Receptor and DNA Coding Therefor
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036

```
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
;   PATENT IN RELEASE #1.24
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/132.990A
;   FILING DATE: 07-OCT-1993
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/084,729
;   FILING DATE: 29-JUN-1993
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US93/05569
;   FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/899,075
;   FILING DATE: 11-JUN-1992
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/806,178
;   FILING DATE: 13-DEC-1991
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/627,950
;   FILING DATE: 14-DEC-1990
; ATTORNEY/AGENT INFORMATION:
;   NAME: Mirock, S. Leslie
;   REGISTRATION NUMBER: 18,872
;   REFERENCE/DOCKET NUMBER: 8105-004-999
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 790-9090
;   TELEFAX: (212) 869-8864
;   TELEX: 66441 PENNIE
; INFORMATION FOR SEQ ID NO: 8:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 629 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   US-08-132-990A-8

Query Match          5.4%; Score 145; DB 2; Length 629;
Best Local Similarity 19.2%; Pred. No. 4.3e-06;
Matches 102; Conservative 85; Mismatches 186; Indels 158; Gaps 24;

QY 44 TTTIVSLAFVICSNWAGISGSL--QLALLAG-----GPTVLLYGILLSTLVYICIAFS 94
DB 34 TFDLVAL-----GVGSTLGAGVYVLGAVARENAGP-AIVISFLRALASVLGLC 83

QY 95 LAELTSVYPTAGQYHFAISILAPKSNRSISYVCGVLVSLLSWIAIGSSVTMIPAQIIPAL 154
DB 84 YGEFGARVPKTSAYLYSV---TVGELWAFITGWNLLSYIIGTSSVARAWSATFDEL 139

QY 155 IAAYSHTYSQDSWHVFLIYEG-----VALVVLNLFAL--KRNPPWHEIGFGLT 202
DB 140 IGRPIGEFERT--HMTLNAPGVLAENPDFAVIIILITGLTLGLVKESAMWNKI--FTC 195

QY 203 IALFVIFSIATLARSNPKNAPNSQVMTAWSNYTCWSDGVCF----- 242
DB 196 INVVLGFTMVSGFVKGVKNQQL--TEEDFGNTSGRLCNDNTRKRGKPGVGEMPFPGFS 253

QY 243 --ILGLSTSCFPIGLDAAHMLAEECTDAARTVPKAVSAIIGCTAP----PTTIAVL 296
DB 254 GVLGSAATCFYAFVGDFDCIATTGEEVKNPQKAIPGVIVASLLICFIAYFGVSAALTLMMP 313

QY 297 YGITDLDLSLSA---GY--IPFETMTQSLRSLSFATVLSGGIVMAFFALNAVOETAS 350
DB 314 YFCLDNNSPDPAKFHVHGEGAKYAVAVGSLCAL-----ASLLGSMFPM-----P 359

QY 351 RLTSWFARDNGLVFSHTLERIHPHQVFWSLFATWGILATGCCIF-----LGSS 400
DB 360 RVYIYMAED-GLLFK-FLANVDRNTKTPLIATLASCAGAAVMAFLFDLKDLDLMSIGTL 417

QY 401 TAFNALVANSVVLQQLSFLPIALLLYQKRDPK----- 433
```

```
DB 418 LAYSILVAACVILVR-----YOPEQNLYYQMASTSDDELDPADONELASTNDSQ 465
QY 434 --FLPSTRAFLV-----PRGI-----GFLVN-----VLAVFTSVTVV 464
DB 466 LGFLPEAEAMFSLKTLSPKMEPSKISGLIVNISTSLIAVLITTCIVTVL 516

RESULT 4
PCT-US92-09382-8
; Sequence 8, Application PC/TUS9209382
; GENERAL INFORMATION:
; APPLICANT: MERUELO, DANIEL
; APPLICANT: YOSHIMOTO, TAKAYUKI
; TITLE OF INVENTION: Human Retrovirus Receptor and DNA Coding
; TITLE OF INVENTION: Therefor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/09382
; FILING DATE: 19921213
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Livnat, Shmuel
; REGISTRATION NUMBER: 33,949
; REFERENCE/DOCKET NUMBER: MERUELO-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 628-5197
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 629 amino acids
;   TYPE: AMINO ACID
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   PCT-US92-09382-8

Query Match          5.4%; Score 145; DB 5; Length 629;
Best Local Similarity 19.2%; Pred. No. 4.3e-06;
Matches 102; Conservative 85; Mismatches 186; Indels 158; Gaps 24;

QY 44 TTTIVSLAFVICSNWAGISGSL--QLALLAG-----GPTVLLYGILLSTLVYICIAFS 94
DB 34 TFDLVAL-----GVGSTLGAGVYVLGAVARENAGP-AIVISFLRALASVLGLC 83

QY 95 LAELTSVYPTAGQYHFAISILAPKSNRSISYVCGVLVSLLSWIAIGSSVTMIPAQIIPAL 154
DB 84 YGEFGARVPKTSAYLYSV---TVGELWAFITGWNLLSYIIGTSSVARAWSATFDEL 139

QY 155 IAAYSHTYSQDSWHVFLIYEG-----VALVVLNLFAL--KRNPPWHEIGFGLT 202
DB 140 IGRPIGEFERT--HMTLNAPGVLAENPDFAVIIILITGLTLGLVKESAMWNKI--FTC 195

QY 203 IALFVIFSIATLARSNPKNAPNSQVMTAWSNYTCWSDGVCF----- 242
DB 196 INVVLGFTMVSGFVKGVKNQQL--TEEDFGNTSGRLCNDNTRKRGKPGVGEMPFPGFS 253

QY 243 --ILGLSTSCFPIGLDAAHMLAEECTDAARTVPKAVSAIIGCTAP----PTTIAVL 296
DB 254 GVLGSAATCFYAFVGDFDCIATTGEEVKNPQKAIPGVIVASLLICFIAYFGVSAALTLMMP 313

QY 297 YGITDLDLSLSA---GY--IPFETMTQSLRSLSFATVLSGGIVMAFFALNAVOETAS 350
```

Db 314 YFCDDNNSPLDPAKFKHVGWEGAKYAVAVGSLCALSSLLGSMFPM-----P 359
Qy 351 RLWTFARDNGLVFTSLHLERHPRWQVWMSLFATWGLATCGCIF-----LGSS 400
Db 360 RVIVAMARD-CULFK-FLANVDRKTPITATLASGAVAAWMAFLDKLDVLDLMSIGTL 417
Qy 401 TAFNALVNSAVVLOOLSLPIALLLYOKRDPK-----433
Db 418 LAYSLVAACVLVLR-----YQEPQNLVYQMASTSDLEDPADQNELASTNDSQ 465
Qy 434 --FLPSTRAFVL-----PRGI-----GFLVN-----VLAVVFTSVTV 464
Db 466 LGFLPEREMFSLKTLSPKKNMPSKISGLIVNIISTLSIAVLIITFCIVTVL 516

RESULT 5
US-09-073-362-1
; Sequence 1, Application US/09073362
; Patent No. 5942399
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: AMINO ACID PERMEASE HOMOLOG
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,362
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Cerrone, Michael C
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0514 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 511 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: ESOCT02
; CLONE: 2667831
US-09-073-362-1

Query Match 5.1%; Score 137.5; DB 2; Length 511;
Best Local Similarity 21.3%; Pred. No. 1.9e-05;
Matches 119; Conservative 90; Mismatches 216; Indels 133; Gaps 28;
Qy 1 MDSRPSGKGGKGRQTKNTETAAAGGASLNVLPKKGFGFTITIVSLAFVICNSWAG 60
Db 15 VETSPLDGSGPGPEQVKKKEISLLNGVCLIVG-----NMIGSGIFVSPKGVLIYS--- 66

Qy 61 ISGSLQIALLAGGPVTLTYGILISTLVVICIAFSLAELTSVPTAGGYHFPASILAPKSI 120
Db 67 ASFGLSLVIWAVGLGFSVFG-----ALCYAELGTTIKKSGASY----- 104
Qy 121 NRSISYVC-GLVSLLSMTAIGSSVTMI-PAOQ-----IPALIAAYSHYYS 163
Db 105 ---AYILEAFGGFLAFIRLWTSLLIETPTSOAILAITFANYMVQPLFPSCFAPYAASR- 159
Qy 164 ODSWHVFLIYEGVALVLLFNLFALKRNPWYHEIGFGTLTIALFVIFIAILARSNPKAPN 223
Db 160 -----LAAACICLLTFINCAVYKMGCTLVODI---FTYAK-VLALIAVIVAGIVRLCQ 208
Qy 224 SOVWTANSNYTGWSDGVCFILG-----LSTSCFMFIGLDAAMHLAEECTDAARTVPKAV- 277
Db 209 G----ASTHFENSEFGSSFAVGDIALALYSALFSTSGMDTLNYVTEETKNPERNPLSIG 264
Qy 278 VSAIIIGFCTAPPTYIAVLYGITDLSILSS-AGYIPETMTQSLRS--LSFATVLSGG 334
Db 265 ISMPIV--TIIVILTNAVYTVLDMRDILASDAVAVTFADQIFGIFNWIIPLSVALSC-- 320
Qy 335 IVMAFFALNAVOETASRLTWSFARDNGL---VFSTHLERIHPRWQVWMSLFATWGLAT 391
Db 321 ----FGGLNASIVAAARLFFVSGRGLPDAICMIVHERF-----TPVPSLLFN-GIMA- 369
Qy 392 CGCIFLGSSTAFNALVNSAVVLOOLSF-----LPIALLLYQKRDPKFLPSTRAFVLP 444
Db 370 --LIYLCVEDIFQ-----LINYYSFSYWEFFVGLSVGQLYLRWKEPD-----RP 411
Qy 445 RGIGFLVNLAVFTSVTVTFSPPLVPTAASMTMNTYSALIGVALGVLNW---VVHA 501
Db 412 RPL-----KLSVFFPIVCLCTIFLVAVPLVSDTIN---SLIGIAIALSGLPFFYLIIRV 463
Qy 502 RKHYGGPHLELDGRVGA 519
Db 464 PEKRPYLRL---RIVGS 478

RESULT 6
US-09-243-920-1
; Sequence 1, Application US/09243920
; Patent No. 5981242
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: AMINO ACID PERMEASE HOMOLOG
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/243,920
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/073,362
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Cerrone, Michael C
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0514 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166


```
TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 511 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: ESOGTUT02
; CLONE: 2667831
;
US-09-243-920-1

Query Match 5.1%; Score 137.5; DB 2; Length 511;
Best Local Similarity 21.3%; Pred. No. 1.8e-05;
Matches 119; Conservative 90; Mismatches 216; Indels 133; Gaps 28;

QY 1 MDSRPSGKGGTGTTRNTTAAAGGASESLNVPLEKKQFGTITIVSLAFVICSNWAG 60
Db 15 VETSPGDSAGPQGVKLUKEISLLNGVCLIVG-----NMIGSGIFVSPKGVLIYS--- 66
QY 61 IGSGLQALLAGGPVLLYGLISTLVYICIAFSLAELTSVYPTAGQYHFASILAPKSI 120
Db 67 ASFGLSLVIWVGGFLSVFG-----ALCYAELGTTIKKSGASY----- 104
QY 121 NRSISYVC-GLYSLLSWIAIGSSVTMI-PAQQ-----IPALIAAYSHYYS 163
Db 105 ----ATILEAFGGFLAFIRLWTSLLIETPSQAIITAFNTYMWQPLFPSCFAPYAAASR- 159
QY 164 ODSWHVFLIYEGVALVLLFNLFALKRNPWVHEIGFGLTIALFVISTAILARSNPKAPN 223
Db 160 -----LLAAACICLLTTFINCAVYKVGTLVQDI---FTYAK-VLALIAVIVAGIVRLQO 208
QY 224 SQWTAWSNYTGSDGVCILG-----LSTSCFMIGLDAAMHLAECTDAARTVPKAV- 277
Db 209 G-----ASTHFENSFGSSPAVGDIALALYSALFSYSGWDTLNYVTEETKNPERNPLSLIG 264
QY 278 VSAIIGFCTAPYITAVLYGITDLSLSS-AGVIPPETMTQSLRS--LSFATVLSCGG 334
Db 265 ISMPIV--TIYILTNAVYTVLDRDILASDAVAVTADQIFGFNWIPLSVLSC-- 320
QY 335 IVMAFALNAOERTASRLTWSFARNGL---VFSTHLERIHRPWOVPMVWSLAFATWGILAT 391
Db 321 ----FGLNASIAASRLRFFVSGREGHLPOAICMTHVERF-----TPVPSLLFN-GIMA- 369
QY 392 CGCIFLGSSTAFNALVNSAVLQQLSF-----LPIATLLYQKRDPKFLPSTRAFLVP 444
Db 370 --LIVLCVEDIFQ-----LINYSFSYFWFVGLSVIGQLYLRKEPD-----RP 411
QY 445 RGIGFLVNLVAVFTSVTVTPFSFPLTVPTAASTNMNYTSIIGVALGVLNM---VVHA 501
Db 412 RPL-----KLSVFFFPVFCITFLVAVPLYSDTIN--SLIGIAIALSGLPFYFLIIRV 463
QY 502 RKHYOGPHLELDGRVVGA 519
Db 464 PEKKRPLYLR---RIVGS 478

RESULT 7
US-09-073-362-3
; Sequence 3, Application US/09073362
; Patent No. 5942399
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: AMINO ACID PERMEASE HOMOLOG
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
```

```
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073,362
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Certone, Michael C
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0514 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-845-0555
TELEFAX: 650-845-4166
TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1665759
;
US-09-073-362-3

Query Match 4.8%; Score 129.5; DB 2; Length 515;
Best Local Similarity 22.0%; Pred. No. 0.00012;
Matches 112; Conservative 77; Mismatches 192; Indels 127; Gaps 27;

QY 29 ASESLNVPLEKKQFGTITIVSLAFVICSNWAGISLQALLAGGPVLLYGLISTLVY 88
Db 35 SSETMQL---KKEISLLNGVSL--VVGN-----MIGSGIFVSPKGVLVHTASY 77
QY 89 -----ICIAFSLAELTSVYPTAGQYHFASILAPKSIINRSISYVCGLVSL 134
Db 78 GMSLIVWAIGGLFSVVGALCYAELGTTIKSGASY--AVIL--EAFGGPIAFIRLWVSL 133
QY 135 SWIATGSSVTMI-----PAQQIPALIAAYSHYYSQDSWHVFLIYEGVALVLLF 183
Db 134 VVEPTGOAIIATFANYIIQPSFPCDPPYLACR-----LAAACICLLTFV 180
QY 184 NLFALKRNPWVHEIGFGLTIALFVISTAILARSNPK--APNSQVMTAWSNTGSDGVC 241
Db 181 NCAYVK---WGRVODTFTYAK--VVALIAIIVMGLVKLCQGHSEHFQDAFEGSSWDMG-N 235
QY 242 FILGLSTCFMFIGLDAAMHLAECTDAARTVPKAV--VSAIIGFCTAPPYITAVLYGIT 300
Db 236 LSLALYSALFSYSGWDTLNFVTEETKNPERNPLAIGSMPIV--TLIYILTNAVYTVL 293
QY 301 LLDLSILSS-AGVIPPETMTQSLRS--LSFATVLSC--GGIVMAFFALNAVQETASRLTWSF 356
Db 294 NISDVLSSDAVAVTADQTFGFMFSWTIPIAVALSCFGLNASIFA-----SSRLFFVG 346
QY 357 ARDN---GLVSTHLERIHRPWOVPMVWSLAFATWGILATC--GCIFLGSSTAFNALVNSAV 411
Db 347 SREGHLPLLSSMIHIERFTp---IP-----ALLFNCTMALIYLIVEDVFQ----- 388
QY 412 VLOQLSF-----LPIATLLYQKRDPKFLPSTRAFLVPRGIGFLVNLVAVFTSVTV 464
Db 389 LINYSFSYFWFVGLSVIGQLYLRKEPK-----RPRPL-----KLSVFFFPVFCI 434
QY 465 FFSFPLTVPTAASTNMNYTSIIGVALAL 492
Db 435 CSVFLVIVPLFTDTIN---SLIGIGIAL 459
```


QY 340 ---FALNAVOETASRLTWSFARDNGLVFSTHLERHPRW-----QVPV-----WSLF 383
Db 287 ALMIAFGVMEVSS---WVGPDSRG-FAAARGLLPKFLRKTNTHEVPVPLVMIOGIIV 342
QY 384 ATWGILATGCCIFLGSSTAFNALVNSAVLQ---QLSFLIPTALLIXQRDPKFLPSTRA 440
Db 343 TLMGAVLTFG---GGGNLSFLVAISUTVVYLVGYLLFFIVFVLYKKONLK-----RT 395
QY 441 FVLP-----RGIGFVLNVLAVVFTSVTTVFFSFPLTVPATASTMYNT 482
Db 396 YNVPGLIKTIAGIGFLLSIFA-----LFISF---VPPASIAKNET 435
RESULT 10
US-08-132-990A-6
; Sequence 6, Application US/08132990A
; Patent No. 5834589
; GENERAL INFORMATION:
; APPLICANT: MERUELO, DANIEL
; APPLICANT: YOSHIMOTO, TAKAYUKI
; TITLE OF INVENTION: Human Retrovirus Receptor and DNA Coding Therefor
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/132,990A
; FILING DATE: 07-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/084,729
; FILING DATE: 29-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/05569
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/899,075
; FILING DATE: 11-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/806,178
; FILING DATE: 13-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/627,950
; FILING DATE: 14-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 8105-004-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66441 PENNIE
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 453 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-132-990A-6

Query Match 4.3%; Score 116.5; Db 2; Length 453;
Best Local Similarity 21.9%; Pred. No. 0.002;
Matches 71; Conservative 55; Mismatches 119; Indels 79; Gaps 17;

QY 211 IATLARSNPKAPNSQVMTAWS-----NYTGWSDGVCFILGLSTSCFMFIGLDAAMHLAECC 266
Db 23 ISASAREPSENGTSIYGAGGPMYGTG-----TLGAAATCFYAFVGFDCIATTTGEEV 76
QY 267 TDAARTVPKAVYSAILIGFCTAF-----PYTIAVLYGTTDLSDLSLSA-----GYIPFET 316
Db 77 RNPQKAIPGIVTSLVCEMAYFVGSAAULTLMPYLYLLDEKSPVAFVVRMGPKAYV 136
QY 317 MTOSLRSLSPATVLSGGIVMAFFALNAVOETASRLTWSFARDNGLVFSTHLERHPRMO 376
Db 137 AGSICALSTS-----LLGSIFPM-----PRVIYMAED-GLLFKC-LAQINSKYK 180
QY 377 VPWWSLFATWGILATGCCIFLGSSTAFNALVN---SAVVILOOLSFLIPIALLL-YQK---- 429
Db 181 TPVIATLSGGAVAAMVAFLF-----DLKALVDMMSGTILMAYSLVAACVLILRYQPLGCV 235
QY 430 ROPKFLP-----STRAFLVPRGIGFLNVLAVVFTSVTTVFFSFPLVPTA 475
Db 236 EQPKYTPKETLESCTNATLKSESQVTMLQGGF-----SURLTF--SPSALPTR 283
QY 476 ASTMNTYSAILG-----VALALGVL 495
Db 284 QSA-SLVSFLVGFALFLILGLSIL 306
RESULT 11
PCT-US92-09382-6
; Sequence 6, Application PC/TUS9209382
; GENERAL INFORMATION:
; APPLICANT: MERUELO, DANIEL
; APPLICANT: YOSHIMOTO, TAKAYUKI
; TITLE OF INVENTION: Human Retrovirus Receptor and DNA Coding
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/09382
; FILING DATE: 19921213
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Livnat, Shmuel
; REGISTRATION NUMBER: 33,949
; REFERENCE/DOCKET NUMBER: MERUELO=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 628-5197
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 453 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US92-09382-6

Query Match 4.3%; Score 116.5; DB 5; Length 453;
Best Local Similarity 21.9%; Pred. No. 0.002;
Matches 71; Conservative 55; Mismatches 119; Indels 79; Gaps 17;
QY 211 IATLARSNPKAPNSQVMTAWS-----NYTGWSDGVCFILGLSTSCFMFIGLDAAMHLAECC 266

Db 23 ISASAREPPSENGSTIYGAGGMPYGFCTG-----TLACAATCFYAFVGFDCIATTGEEV 76
QY 267 TDAARTVPAVSAIIIGCTAF-----PYTIAVLYGITDLSILSSA-----GYIPFET 316
Db 77 RNPQRAIPGIVTSLVCFMAYFGVSAALTLMPPYLLDEKSPLPVAFEFVRWGPAYV 136
QY 317 MTQSLRSLSFATVLSGGGLVMAFFALNAVOETASRLTWSFARNGLVFSTHLERHPRWQ 376
Db 137 AAGSLCALSTS-----LLGSIFPM-----PRVIYAAED-GLLFKC-LAQINSKTK 180
QY 377 VPVWSLFWATGILATCGCIFLGSSTAFNALVN--SAVVLQOOLSFLPIALLL-YOK-----429
Db 181 TPVIATLSGAVAAVMAFLF-----DLKALVDMSTGTLMAYSLAACVLAAILRYOPGLCY 235
QY 430 RDPKFLP-----STRAFVLPGRIGFLVNLAVFTSVTTVFESPLPVPFTA 475
Db 236 EQPKVTPKETLESCTNATLKSESQVTMLQOGF-----SLRTLF--SPSALPTR 283
QY 476 ASTMYTSAIIG-----VALALGVL 495
Db 284 QSA-SLVSLVGLFLAFLILGLSIL 306

RESULT 12
US-08-362-512A-4
; Sequence 4, Application US/08362512A
; Patent No. 5719043
; GENERAL INFORMATION:
; APPLICANT: FROMMER, Wolf-Bernd
; TITLE OF INVENTION: DNA SEQUENCES FOR AN AMINO ACID
; TITLE OF INVENTION: TRANSPORTER, PLASMIDS, BACTERIA, YEASTS AND PLANTS
; TITLE OF INVENTION: CONTAINING A TRANSPORTER AND THEIR USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362.512A
; FILING DATE: 05-JAN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/007,636
; FILING DATE: 21-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meilman, Edward A.
; REGISTRATION NUMBER: 24,735
; REFERENCE/DOCKET NUMBER: P/951-107
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 382-0700
; TELEFAX: (212) 382-0888
; TELEX: 236925
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 493 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-362-512A-4

Query Match 4.2%; Score 114; DB 1; Length 493;
Best Local Similarity 21.9%; Pred. No. 0.0041;
Matches 111; Conservative 72; Mismatches 149; Indels 174; Gaps 32;

QY 40 KQGT-----TTIVSLAFVQICNSWAGISGSLQALLAGGPVTLXYGILSTLVYICI 91
Db 45 KRGTGVTWASAHIITAVIGSGVLSLAWA-----IAQLGWIAGPAVMLFSLV--TL-----93
QY 92 AFSLAELTSVYPT---AGGYHFASILAPKINSISY-VCGLVSLLS--WIAIGSSVTM 145
Db 94 -YSTLSLSDCYRTGDVSGKRNYYTMDAVRSILGGFKFKICGLIQYLNLFGLGIAGYTI--150
QY 146 IPAQOIPALIAAYSHTY-----SODSMWFLIVEGVALVLLFNLFALKRNPVWHEIGFGL 201
Db 151 --RASIISMAIKRSCNCFHSGGKDPCH-----MSSNPYM--IVFG--186
QY 202 TIALFVIFSTAILARSNPKAPN-SQVMTAMSNYTGWSDGVCYFILGLSTSCFMFIGLDAAM 260
Db 187 -VAEILLS-----QVPDFDOIW-----WISIVAAMVSFTYSA---IGL--AL 222
QY 261 HLAECTDAARTVPKAVVSAILIGFCT-----APPYTTAVLYGITDLSI 305
Db 223 GIVQV-----AANGVPKGLTGISIGTVTQTKIMRTFOALGDIAFAYSVVL-IEIQDTV 278
QY 306 LSSAGYIPFE--TMTQSLR-SLSPATV--LSCGGIVMAFFALNAVOETASRLTWSFARDN 360
Db 279 RSP-----PAESKTMKKATKISIAVTTFYMLCGSMGYAAGF---DAAPGNLLTGGGFYN 330
QY 361 GLVFSTHLERHPRWQVPMVWSLFPATWGILATCGCIFLGSSTAFNALVNSAVVLOOLS---417
Db 331 -----PFWLL-----DIANNAIIVHLVGAYQ 351
QY 418 -FLIPTALLYOK-----RDPKFLPSTRAFVLPGRIGF---LVNLAVVFTSVTTVFFS 467
Db 352 VFAQTPAFATEKSAERYPNDFLSKEFEIRIP---GFKSPYKVNVRMVRYS-----G 402
QY 468 FLPTVPTAASTMYTSAIIGVALG 493
Db 403 FVVTTVISMPLMPFENDVVGILGALG 428

RESULT 13
US-08-964-939-4
; Sequence 4, Application US/08964939
; Patent No. 6245970
; GENERAL INFORMATION:
; APPLICANT: FROMMER, Wolf-Bernd
; TITLE OF INVENTION: DNA SEQUENCES FOR AN AMINO ACID
; TITLE OF INVENTION: TRANSPORTER, PLASMIDS, BACTERIA, YEASTS AND PLANTS
; TITLE OF INVENTION: CONTAINING A TRANSPORTER AND THEIR USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/964,939
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,512
; FILING DATE: 05-JAN-1995
; APPLICATION NUMBER: US 08/007,636
; FILING DATE: 21-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meilman, Edward A.
; REGISTRATION NUMBER: 24,735
; REFERENCE/DOCKET NUMBER: P/951-107
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 382-0700
TELEFAX: (212) 382-0888
TELEX: 236925
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 493 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-964-939-4

Query Match 4.2%; Score 114; DB 4; Length 493;
Best Local Similarity 21.9%; Pred. No. 0.0041;
Matches 111; Conservative 72; Mismatches 149; Indels 174; Gaps 32;
QY 40 KQGT-----ITTVSLAFVICSNWAGISGLQALLAGGPTVLLYGLISTLVYICI 91
DB 45 KRGTWTASAHITAVIGSVLSLARA-----IAQLGWAGPVMVLLFSLV--TL----- 93
QY 92 AFSLAELTSVYPT---AGQYHFASILAPKSNRSISY-VCGLVSLLS--WIAIGSSVTM 145
DB 94 -YSSTLLDCYRTGDVAGSKRNYTMDVRSILGGFKKICGLIQVNLFGIAIGVTI-- 150
QY 146 IPAQOIPALIAAYSHY-----SODSHVFLIYEGVALVLLNLFALKRNPVWHEIGFGL 201
DB 151 --AASISMAIKRSCFHKSGKDPCH-----MSSNPYM--IVFG- 186
QY 202 TIALFVIFSIATILARSNKPKN--SQWTAWSNYTGWSDGVCFLGLSTSCFMFIGLDAAM 260
DB 187 -VAEILLS-----QVPDQIWI-----WISIVAAVMSPTYSA---IGL--AL 222
QY 261 HLAECTDAARTVPKAVSAIIIGCT-----AFPTIATVIGITDLSI 305
DB 223 GIVGV---AANGVKRSLTGISGTGTQKWTWTFQALGDIAFAYSYSWVL-IEQDTV 278
QY 306 LGSAGYIPPE--TMTQSLR-SLSFAFV--LSCGGIVMAFFALNAVQETASRLTWSFARN 360
DB 279 RSP---PAESKTKKATISIAVTITFVMLCGSNCGYAAFG---DAAPGNLLTGFGFN 330
QY 361 GLVFSTHLERIPRWQVPVWSLENTWGLATCGCIFLGSSTAFNALVNSAVVLOQLS--- 417
DB 331 -----PFMLL-----DIANAIAVHVHLGAYQ 351
QY 418 -FLIPIALLLYOK-----RDKPFLPSTRAFLVPRGIGF-----LVNVLAUVFTSVTTVPFS 467
DB 352 VFAQPIFAIEKSVAEYRDNDFLSKEFEIRIP--GFKSPYKVVNVFRMYRS-----G 402
QY 468 FPLTVPTAASTWNYTSAILGVVALG 493
DB 403 FVYTTTVISMLMPFFNDVVVGILGALG 428

RESULT 14
US-09-097-889-15
Sequence 15, Application US/09097889
Patent No. 6218117
GENERAL INFORMATION:
APPLICANT: Herrnstadt, Corrina
APPLICANT: Ghosh, Soumitra S.
APPLICANT: Davis, Robert E.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING
TITLE OF INVENTION: AGENTS THAT QUANTITATIVELY ALTER DETECTABLE
TITLE OF INVENTION: EXTRAMITOCHONDRIAL DNA: MITOCHONDRIAL DNA RATIOS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/097,889
FILING DATE: 15-JUN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Roseman Ph.D., Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 660088.417
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 513 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-097-889-15

Query Match 4.1%; Score 109.5; DB 4; Length 513;
Best Local Similarity 17.3%; Pred. No. 0.012;
Matches 87; Conservative 69; Mismatches 139; Indels 209; Gaps 24;
QY 40 KQGTITVSLAFVICSNWAGISGLQALLAGGPTVLL-----YGLISTLVYICI 91
DB 13 KDIGIL-----YLLFGAWAGVLGTALSLIRAEGLQPGNLLGNHIIYVITAHAFVMI 66
QY 92 AFSLAELT----- 99
DB 67 FTWVMPIMIGFGNWLVLPLMIGAPDMAFPRMNMNSFWLLPPLLLLLASAMVEAGATGW 126
QY 100 SVYPTAGGY-----HFASILAPKSNRSISYVCGLV-----L 133
DB 127 TVYPPLAGNYSHPGASVDLTIESLHAGV---SSILGAINFITTIINMKPPANTQYOTPL 183
QY 134 LSWIAGTSVTMPAQOIPALIAA-----YSHTYSODSWH 168
DB 184 FVMSVLITAVLL--LSLPVLAAGITMLLTDRLNLTFFDPAGGGDPILYQHLF--WF 237
QY 169 -----VFLIYEGVALVLLNLFALKRNPVWHEIGFGLTIALFVLSFTAILARSNKPAP 222
DB 238 FGHPEVYIILPGFMISHIYVYSGKKEPGY---MGVWAMMSIGLGF----- 286
QY 223 NSQVWTAMSNYTGWSGVCFLGLS-----TSCFMFIGLD-----AAMHLAEC 266
DB 287 -----VMAHH-----MFTVMDVDTRAYFTSATMIIAIPTKVKSFWLATLHGSNMK 333
QY 267 TDAARTVPKAVSAIIIGCTAPPYTIATVLYCI-----TDLDSILSSAGYIPETWTOSLR 322
DB 334 WSA-----AVLWA--LGP--IFLFTVGLGTGIVLANSSLDIVLHDTYIV----- 373
QY 323 SLSFATVLSGGIV-----MAFFALNAVQETASRL--TWSEFARDNGLVFSTHLERI 371
DB 374 VAHFHVYLSMGAVFALMGGFIFHWFLPSGYTLDTQIVAKIHTIMEIGVNLTFPQHFLGL 433
QY 372 H--PRQVPMVNSLFWATWGLATCG 393
DB 434 SGMPRYSYDPAATYTNWILSSVG 457

RESULT 15
US-08-677-049-9
Sequence 9, Application US/08677049
Patent No. 5858707
GENERAL INFORMATION:
APPLICANT: Guimaraes, M. Jorge
APPLICANT: Bazan, J. Fernando
APPLICANT: McClanahan, Terrill K.

```

1  APPLICANT:  Zlotnik, Albert
2  TITLE OF INVENTION:  PURIFIED MAMMALIAN NUCLEOBASE PERMEASES;
3  TITLE OF INVENTION:  NUCLEIC ACIDS; ANTIBODIES
4  NUMBER OF SEQUENCES:  12
5  CORRESPONDENCE ADDRESS:
6
7  ADDRESSEE:  DNAX Research Institute
8  STREET:  901 California Avenue
9  CITY:  Palo Alto
10 STATE:  California
11 COUNTRY:  USA
12
13 ZIP:  94304-1104
14
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE:  Floppy disk
17 COMPUTER:  IBM PC compatible
18 OPERATING SYSTEM:  PC-DOS/MS-DOS
19 SOFTWARE:  PatentIn Release #1.0, Version #1.30
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER:  US/08/677,049
22 FILING DATE:  03-JUL-1996
23 CLASSIFICATION:  435
24
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER:  US 60/000,788
27 FILING DATE:  03-JUL-1996
28 ATTORNEY/AGENT INFORMATION:
29 NAME:  Ching, Edwin P.
30 REGISTRATION NUMBER:  34,090
31 REFERENCE/DOCKET NUMBER:  DX0511
32 TELECOMMUNICATION INFORMATION:
33 TELEPHONE:  415-852-9196
34 TELEFAX:  415-496-1200
35 INFORMATION FOR SEQ ID NO:  9:
36 SEQUENCE CHARACTERISTICS:
37 LENGTH:  438 amino acids
38 TYPE:  amino acid
39 STRANDEDNESS:  single
40 TOPOLOGY:  linear
41 MOLECULE TYPE:  protein
42 FEATURE:
43 NAME/KEY:  Region
44 LOCATION:  115..144
45 OTHER INFORMATION:
46 OTHER INFORMATION:  4"
47
48 FEATURE:
49 NAME/KEY:  Region
50 LOCATION:  323..357
51 OTHER INFORMATION:
52 OTHER INFORMATION:  4"
53
54 FEATURE:
55 NAME/KEY:  Region
56 LOCATION:  359..386
57 OTHER INFORMATION:
58 OTHER INFORMATION:  4"
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934

```

	Query Match	4.0%	Score 107;	DB 2:	Length 438;
	Best Local Similarity	19.8%;	Pred.	No. 0.018;	
	Matches 102; Conservative	78;	Mismatches	172;	Indels 164; Gaps 25;
QY	42 FGRTITVSLAFVICNSWAGISGSQLALLAGGPVTLL---	YGILISTLVYICIAFSLAEL	98		
Dd	: :	: :	:	:	:
	5 FKGTLGLGIQHVL----	AMVAGAIVPELIIVKMGMLTVEQLTYLVSIDIFMCGVATLLQV	60		
QY	99 -----	TSPVP--TAGGOYH-----	FASIIAPKSINRSIYVCVG-LVSL	133	
Dd					
	61 WSNRFFGIGLPVVLCGTFTAVSWMIAIGSYGVSTVGSITIASGILLWILISFFFGKKLVSF	120			
QY	134 LSWNIATGSSVTMTPAQOIPALIRAYSHITYSQDSWHHVFPIEGVALVLFNPLFAALKRNPW	193			
Dd	:	: :	:	:	:
	121 FPVPVTVGSVTTI-----	GITLMPVAMNMNMAAGGE-	-- 150		
QY	194 VHEIGFG-----LTIALFVISFIAILARSNPKNPSQVMWTAWMSNYTGSDGVCFFTLIGLSTS	249			
	: :	: :	:	:	:

Db	151	--SADFGLSNLAFTVLSIIIVLLYRFT-----KGFKSVSILIGILIG 193
Qy	250	CFM--FIG-----LDAAM-----HLAEECTDAARTVPKAVWSAIIIGCTAFPYTI 293
Db	194	TFIATYFGKQVQDNVSDAAVQMIQPFYFCAPSFHAAPITMSIVALSIVESTGYVFAL 253
Qy	294	AVLYG--ITDLDSTLSNAGYIPETWTQSLRSLSFATVLSGCGIVMAFPALNAVQBTASR 351
Db	254	GDLNRRRLTEID---LSKGY-----RAGGLAVLL--GGI-----FNAPFYTA-- 290
Qy	352	LTWSEFARDNGILVSTHIERIHRMQVPWSLFTAGTGCIGIFLG-----SSTAFNAL 406
Db	291	---FSQNVGLVQLTGIKKN-----AVIVVTGVILMAFLGFPKIAAFTTI 331
Qy	407	VNSAV-----VLOQLSFLIPALTLLYQKRDPKPLPSTPAFLPRGICFLVNLAVVTSVT 462
Db	332	IPSAVLGGAMVPGMWIAVGIKMLSRID--FAKQENLLIIVACSVG-----LGLGVTVVP 384
Qy	463	TVFFSFELTVPТАASTNNYTSIAIGVALALGVLNW 498
Db	385	DIFKOLP-----SALTLLTNGIVAGSFVAVLNIV 415

Search completed: April 27, 2002, 07:46:48
Job time: 6705 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:45:33 ; Search time 104.96 Seconds
(without alignments)
555.581 Million cell updates/sec

Title: US-09-882-694A-8
Perfect score: 2692
Sequence: 1 MDSRPSGYGKGTQTKN.....QGPHELDGRVVGAEFQVGP 525

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneqeq-embl/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneqeq-embl/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneqeq-embl/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneqeq-embl/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneqeq-embl/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneqeq-embl/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneqeq-embl/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneqeq-embl/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneqeq-embl/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneqeq-embl/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneqeq-embl/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneqeq-embl/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneqeq-embl/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneqeq-embl/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneqeq-embl/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneqeq-embl/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneqeq-embl/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneqeq-embl/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneqeq-embl/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneqeq-embl/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneqeq-embl/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneqeq-embl/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	98.6	2655	525	AA1980	Exophiala spinifera
2	451	16.8	505	AA1981	Arabidopsis thaliana
3	451	16.8	517	AA1982	Arabidopsis thaliana
4	443.5	16.7	504	AA1983	Arabidopsis thaliana
5	449.5	16.7	504	AA1984	Arabidopsis thaliana
6	449.5	16.7	516	AA1985	Arabidopsis thaliana
7	449.5	16.7	516	AA1986	Arabidopsis thaliana
8	449.5	16.7	528	AA1987	Arabidopsis thaliana
9	400	14.9	439	AA1988	Arabidopsis thaliana
10	398.5	14.8	437	AA1989	Arabidopsis thaliana
11	326.5	12.1	439	AA1990	Arabidopsis thaliana

12	326.5	12.1	521	21	AA19910	Arabidopsis thaliana
13	305	11.3	311	21	AA19912	Arabidopsis thaliana
14	216	8.0	482	22	AA19913	Staphylococcus aureus
15	204.5	7.6	453	22	AA19914	C. glutamicum protease
16	197	7.3	469	22	AA19915	Propionibacterium
17	192	7.1	412	22	AA19916	S. epidermidis ope
18	190	7.1	456	22	AA19917	Enterococcus faecalis
19	190	7.1	463	22	AA19918	Enterococcus faecalis
20	182	6.8	508	22	AA19919	Propionibacterium
21	179.5	6.7	475	22	AA19920	Pseudomonas aeruginosa
22	176	6.5	530	22	AA19921	Amino acid sequence
23	175	6.5	499	22	AA19922	Drosophila melanogaster
24	173	6.4	508	22	AA19923	Propionibacterium
25	171	6.4	466	22	AA19924	Salmonella typhimurium
26	169.5	6.3	482	22	AA19925	E. coli growth and
27	169	6.3	390	22	AA19926	S. epidermidis ope
28	167	6.2	544	22	AA19927	Propionibacterium
29	164.5	6.1	489	22	AA19928	E. coli cellular p
30	164.5	6.1	489	22	AA19929	Escherichia coli l
31	164	6.1	523	22	AA19930	Amino acid sequence
32	164	6.1	622	13	AA19931	ERR receptor. Mus
33	164	6.1	622	20	AA19932	Murine ecotropic r
34	163.5	6.1	804	22	AA19933	Drosophila melanogaster
35	162	6.0	466	22	AA19934	Propionibacterium
36	162	6.0	540	22	AA19935	E. coli cellular p
37	161	6.0	466	14	AA19936	Sequence of gamma-
38	160	5.9	880	22	AA19937	Drosophila melanogaster
39	159	5.9	489	22	AA19938	Salmonella typhimurium
40	159	5.9	510	21	AA19939	Arabidopsis thaliana
41	159	5.9	585	21	AA19940	Arabidopsis thaliana
42	159	5.9	590	21	AA19941	Arabidopsis thaliana
43	158.5	5.9	455	22	AA19942	Propionibacterium
44	157.5	5.9	533	22	AA19943	Rat L-type amino a
45	157.5	5.9	604	22	AA19944	Drosophila melanogaster

ALIGNMENTS

RESULT 1
AA19944
ID AA19944 standard; Protein: 525 AA.
AC AA19944
CC AA19944
DT 02-MAY-2001 (first entry)
XX Exophiala spinifera permealase, a funonisin degradative enzyme.
DE Fumonisin; degradative enzyme; transport enzyme; flavin monooxygenase; aldehyde dehydrogenase; permease; p-glycoprotein; fumonisin esterase; detoxification; mycotoxin; animal feed; human feed; silage; transgenic plant; transgenic animal; microbial spray.
XX Exophiala spinifera.
XX Exophiala spinifera.
XX WO200105980-A1.
XX 25-JAN-2001.
XX 14-JUL-1999; 99WO-US15824.
XX 14-JUL-1999; 99WO-US15824.
XX (PION-) PIONEER HI-BRED INT INC.
XX (CURA-) CURAGEN CORP.
XX Duvick JP, Maddox J, Gilliam J, Folkerts O, Crasta OR; WPI; 2001-147345/15.
XX N-PSDB; AAD02693.
XX Novel polynucleotides encoding Exophiala degradative or transport

PT enzyme which is useful for detoxifying fumonisin or structurally
PT related mycotoxin during processing of grain for human or animal food
XX consumption -

XX Claim lc; Page 71-73; 90pp; English.

XX The patent discloses novel polynucleotides encoding *Exophiala spinifera*
CC fumonisin degradative or transport enzymes such as flavin monooxygenase,
CC aldehyde dehydrogenase, permease and p-glycoprotein. A fumonisin esterase
CC enzyme with at least one fumonisin degradative enzyme is useful for
CC detoxifying fumonisin or a structurally related mycotoxin present in
CC harvested grain, on application to a plant or to harvested grain
CC during processing, or to processed grain that is to be used as animal
CC or human feed, or as a silage. Nucleotide fragments of the present
CC invention are useful as probes and primers. They can be introduced
CC into microorganisms that multiply on plants to deliver enzymes to
CC potential target crops. The genes encoding the degrading enzymes are
CC introduced via a vector into a microbial host and the transformed host
CC is supplied to the environment, plants or animals for reducing the
CC pathogenicity of a fungus producing fumonisin. The genes of the
CC invention are fermented in a bacterial host and the resulting bacteria
CC is processed and used as a microbial spray. The nucleotide sequences
CC can be used alone or in combination to engineer microbes or other
CC organisms to metabolise fumonisin and resist its toxic effects.
CC The present protein sequence is permease, a fumonisin degradative
CC enzyme from *Exophiala spinifera*.

XX Sequence 525 AA;

Query Match 98.6%; Score 2655; DB 22; Length 525;
Best Local Similarity 98.9%; Pred. No. 3,9e-367;
Matches 519; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 MDSRPSGVEGKGRPTTKNETAAAGGASESLNVPLEKKQFGTITIVSLAFVICSNSWAG 60
DB 1 mdsrpsgyglk9grtqtknctetaagaseslnvplekqfgititvslafvicsnswag 60
QY 61 TSGSIQLALLAGGPTVTLIGLITLVYICIAFSLABELTSVYPAGQYHFASILAPKSI 120
DB 61 tsgsiqlallaggpvtllylgilistlvyciafslaeltsvypagqyghfasilapksi 120
QY 121 NRSISYVCGLVSLSWTAIGSSVTMIPAQQIPALIAAYSHYTSQDSNHFLLIVEGVALV 180
DB 121 nrsisyvcglvslswtaigssvtmlpaqqipalialaayshytsqdsnvhflyegvalv 180
QY 181 LFLNLFALKRNPWHEIGFGLTIALFVISTAILARSNKAPNSQVMTANSNTGWSGV 240
DB 181 llfnlfalkrnpwheigfgltialfvistailarsnkapnsqvmtansntgwsdgv 240
QY 241 CFIIGLSTSCPMFTGLDAAHMLAECDTDAARTVPKAVVSAILIIGFCTAFPTIAYLYGIT 300
DB 241 cfiiglstscfmgfldaahmlaeecdtaartvpkavvsailiigfctafptiaylygit 300
QY 301 DLDSILSAGYIPETMTQSLSRUSLSPATVLSGCIYMAFFALNAVQTASRLTWSFARDN 360
DB 301 dlclsilsgyipetmtqsrlsruslspatvlsdciymaffalnnavqetasrltwsfardn 360
QY 361 GLVFSTHERLHPWQVPVWSLFTWGLATCGCIFLGSSTAFNALVNSAVVLQQLSFLI 420
DB 361 glvfstherlhpwqvpvwsflftwglatcgciflgsstafnalvnsavvlqqisfl 420
QY 421 PIALLLYKROKPKLPSTRAFLPRGTGFLVNLAVVFTSVTVVFFSFPTVPTAATSMN 480
DB 421 pialllykrodkpklpstrafvlprgtgflvnlavvftsvtvvffsfptvptaastmn 480
QY 481 YTSALIGVALGVLNWNVHARKHYQGFHLELDRVVGAEFQVGP 525
DB 481 ytsaligvalalgvlwnvnharkhyqgfhlldrvvgaeqfvgp 525

RESULT 2
AAG31987

ID AAG31987 standard; Protein; 505 AA.

XX AAG31987;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 38507.

XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 28-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 04-MAY-1999; 99US-0132407.

XX 05-MAY-1999; 99US-0132484.

XX 06-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

XX 07-MAY-1999; 99US-0132487.

XX 11-MAY-1999; 99US-0132863.

XX 14-MAY-1999; 99US-0134256.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134221.

XX 18-MAY-1999; 99US-0134370.

XX 19-MAY-1999; 99US-0134768.

XX 20-MAY-1999; 99US-0134941.

XX 21-MAY-1999; 99US-0135124.

XX 24-MAY-1999; 99US-0135353.

XX 25-MAY-1999; 99US-0135629.

XX 27-MAY-1999; 99US-0136021.

XX 28-MAY-1999; 99US-0136392.

XX 01-JUN-1999; 99US-0136782.

XX 03-JUN-1999; 99US-0137222.

XX 04-JUN-1999; 99US-0137528.

XX 07-JUN-1999; 99US-0137724.

XX 08-JUN-1999; 99US-0138094.

XX 10-JUN-1999; 99US-0138540.

XX 10-JUN-1999; 99US-0138847.

XX 14-JUN-1999; 99US-0139119.

XX 16-JUN-1999; 99US-0139452.

XX 16-JUN-1999; 99US-0139453.

XX 17-JUN-1999; 99US-0139492.

XX 18-JUN-1999; 99US-0139454.

XX 18-JUN-1999; 99US-0139455.

XX 18-JUN-1999; 99US-0139456.

XX 18-JUN-1999; 99US-0139457.

XX 18-JUN-1999; 99US-0139458.

XX 18-JUN-1999; 99US-0139459.

XX 18-JUN-1999; 99US-0139460.

[illegible]

```
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.

PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161820.
PR 28-OCT-1999; 99US-0161821.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 16.8%; Score 451; DB 21; Length 517;
Best Local Similarity 27.2%; Pred. No. 7.le-38;
Matches 135; Conservative 104; Mismatches 217; Indels 40; Gaps 16;

Qy 39 KKQFGTITIVSLAFVICONSWAGISGSLQALLAGGPPVTLVGLIISTVLYICIAFSLAEL 98
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 30 krdlsvfnsfaisiisvltgtttntgtlrggtvtlvvgflagsftmcvlgmaei 89
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 99 TSVYPTAGGOYHFASILAPKINSISYV--CGLVSLLSWIAIGSSVVTMPAQOIPALI- 155
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 90 cssyptaggllywsamlagprwaplasmgckvvvclqw-avtasvdfsalagliqvivl 148
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 156 ----AAYSHYVSQDSWHVFLIYEGVALVVLLENLFAKRNWPVHEIG-----FGLTIALF 206
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 149 lstggrnggykgsdfvvgihgghilfihallnslpivsfvgqaaalwnllgviwimi 208
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 207 VISFTAILARSNPKAPNSQVWTANVTG--WSDGVCFILGLSTCFMFIGLDAAMHLA 263
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 209 lipiv-----sterattkfvtfnfntdnglitsayvifvlgllmsqytltydasahmt 263
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 264 EECTDAARTVPKAVSIAIIIGCTAFYPTIAVLYIGITDLSILS----SAGYIPFETMTQ 319
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 264 eetvdadkngprgilsalgsilfgwyilgisyavtdipslisetnnsqgyaiaefyl 323
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 320 SLRSL--SEAFVLSGGI--VMAFFALNAVQETASRLTWSFARDNGLVFSTHLERHPRW 375
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 324 afknrfsggtgvciglvavavifcgmssvtsnrmayafsrddgampmpelwhkvnvr- 382
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 376 QVPVNSLFATW--GILATCGCI-FLGSSTAPNALVNSAVVLOQLSEFIPTALLLYOKRDP 432
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 383 evpi---navwlsalisfcmaltalsivafqamvsatiglyayaipilrlvtlarn- 438
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 433 KFLPSTRAFLVPRGIGFLVNVLAIVTSTVTFVFPFLTPTAATSMNTYSAIIGVALAL 492
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 439 tfvpq--pfslgk-ygmvgvgwvviwvvtisvlslpvaipitaetinytpvavaglval 495
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 493 GVLNVVHARKHYQGP 508
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 496 tlsywlfsarhwtgtp 511

RESULT 4
AAG20282
ID AAG20282 standard; Protein: 504 AA.
XX
AC AAG20282;
XX
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 22407.
XX
KW protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
```

PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 08-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132408.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.

PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match

Best Local Similarity

Matches 132; Conservative 106; Pred. No. 9.7e-38; Indels 37; Gaps 37

QY	39	KQCFITITIVSLAFVNCNSWAGISGSLQALLAGGPVTLXYCILISTLVVICIAFSLAEI	98
Db	18	krdslsvfnfaisfslisvittgittyneglgfpgtvtlvygwflagsftmcvglsmaei	77
QY	99	TSVPTAGQVHEASILAPKSNRSTSYVCGLVLSLSWIAIGSSWTMPAQOIPALI---	155
Db	78	csyytsggllywysamlagprweplasswmtgwnivggwvatsvdfslsqllqvivils	137
QY	156	--AAYSHYTSODSHVFLIYEGVALVLLFNLFALKRNPWWHEIG-----FGLTIALEVI	208
Db	138	tggrrgggkgsdfvbigbgilfihallnslpsivsfisgglaaawnllgvilvmlil	197
QY	209	SFIAILARNKPNKNSOVWTAWSNYTG--WSDGVCFITLGLSTSCFMEIGLDAAMHLAE	265
Db	198	plv-----sterattkfvfnfndnglgitayifvlgllmsqytlitydasamtee	252
QY	266	CTDAARTVPKAVSALIIGECTAPPTIAVLVYGITDLSILS---SAGYIPFETMTQSL	321
Db	253	tvdadkngprgllisaiglsilfvgwylglsyavtdipsliselnngsgyaaeiaefylaf	312
QY	322	RSU---SFATVLVSCGGI--VMAFTALNAVQETASRLTWSFARDNGILVSTHLEIHPRWQV	377
Db	313	knrfsgtggivclgvavavffcgmsstvtsnrmayafsrddgampsplwhkvnsr-ev	371
QY	378	PWMSLFATW--GILATCGCI-FLGSSTAFNALVNSAVVLOQLSFLPIALLLLYKQKDPKF	434
Db	372	pi---navwlsalisfcmaltsgisvafqanvsiatiglylayaipaillrvrtlarn-tf	427
QY	435	LPSTRAFLVPRGIGFLYNVLAVFTSVTTVTFESFPLTVPTAASTMNTVTSALIGVALAIGV	494
Db	428	vpg---pfslgk-ygmvmvgwvaviwvwtisvlfsplvaypitaelinyupvavaglvaitl	484
QY	495	LNWVWHARKHYQGP	508
Db	485	svwlfsarhwtqtp	498

RESULT	5	
AAG49885	AAG49885	
ID	AAG49885 standard; Protein; 504 AA.	
XX	AC	
XX	AAG49885;	
XX	DT	
XX	18-OCT-2000 (first entry)	
XX	DE	
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 63155.	
XX	DE	
XX	Arabidopsis thaliana.	
XX	OS	
XX	Arabidopsis thaliana.	
XX	PN	
XX	EP1033405-A2.	
XX	PD	
XX	06-SEP-2000.	
XX	PF	
XX	25-FEB-2000; 2000EP-0301439.	
XX	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-01343768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139452.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.

PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147182.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151348.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154179.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159233.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.

PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 16.7%; Score 449.5; DB 21; Length 516;
Best Local Similarity 26.7%; Pred. No. 1e-37;
Matches 132; Conservative 106; Mismatches 219; Indels 37; Gaps 14;

QY 39 KQFGTTIVSLAPVICNSWAGISGSLQALLAGAGPVTLLYLILISTVLVVICIAFLAEL 98
DB 30 krdlsvfnsfaifslsvltgitttyntgtrfggtvtlvygflagsfmcvglsmaei 89
QY 99 TSVYPTAGCOYHPFASILAPKSIINRSISYVCGLSLISWIAIGSSVTWMPAQOIPALI--- 155
DB 90 cssyptsgglyysamlagprwagplasmgtgwnlgvqavtasvdfslaqlivivils 149
QY 156 --AAYSHYTSQDSWHVPLIYEGVALVLLPFLNLFALRNPWVHEIG----FGLTIALFVI 208
DB 150 tggrrnggykgsdfvlgthggilfihallnslpislvsfigqlaalwnllgvlvimll 209
QY 209 SFTAILARSNPKAPNSQVWTANSNYG---WSDGVCFLICLSTSCPFICGLDAAMHAAEE 265
DB 210 plv-----sterattkfvtfnfntdngltsyayifvigllmsqylitgysasahmtee 264
QY 266 CTDAARTPKAVYSAILIIGCTAPPYTIIVLYGITOLDLSILS---SAGYIPFETWTQSL 321
DB 265 tvdadkngprgiisaigisilfgwyilgisayavtdipallsetnngsgyalaefylaf 324
QY 322 RSL--SFATVLSGGI--VMAFFALNAVQETASRLTWSFARDNGLVFSTHLERHPMOV 377
DB 325 knrfsgtggivgcvavavavfvcgmssvtsnrmayafsdgampmsplwhkvnser-ev 383
QY 378 PVMSLFATW--GILATCGCI-FLGSGTAFNALVNSAVVLOOLSLFIPIALLLYQKRDPKF 434
DB 384 pi--navwlsaliscmaitslgsivafqamvsiaitiglyiaypilrlrvtlarn-tf 439
QY 435 LPSTRAFVLPGRIGFLVNLVAVVTSVTTVFFSFPLTVPPTAASTMYNTSAIIGVALALGV 494
DB 440 vpg--pfsigk-ygmvgvgvavlvwvvtisvlfslpvayptaetclnytpvavaglvaitl 496
QY 495 LNWVHARKHYQGP 508
DB 497 sywlfsarhwftgp 510

RESULT 7
AAG49884
ID AAG49884 standard; Protein; 516 AA.
XX
AC AAG49884;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63154.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.

PR	05-MAR-1999;	99US-0123160.	PR	14-JUL-1999;	99US-0143624.
PR	09-MAR-1999;	99US-0123548.	PR	15-JUL-1999;	99US-0144005.
PR	23-MAR-1999;	99US-0125788.	PR	16-JUL-1999;	99US-0144085.
PR	25-MAR-1999;	99US-0126584.	PR	16-JUL-1999;	99US-0144325.
PR	29-MAR-1999;	99US-0126785.	PR	19-JUL-1999;	99US-0144331.
PR	01-APR-1999;	99US-0127462.	PR	19-JUL-1999;	99US-0144332.
PR	06-APR-1999;	99US-0128234.	PR	19-JUL-1999;	99US-0144333.
PR	08-APR-1999;	99US-0128714.	PR	19-JUL-1999;	99US-0144334.
PR	16-APR-1999;	99US-0129845.	PR	19-JUL-1999;	99US-0144335.
PR	19-APR-1999;	99US-0130077.	PR	20-JUL-1999;	99US-0144352.
PR	21-APR-1999;	99US-0130449.	PR	20-JUL-1999;	99US-0144632.
PR	23-APR-1999;	99US-0130510.	PR	20-JUL-1999;	99US-0144632.
PR	23-APR-1999;	99US-0130891.	PR	20-JUL-1999;	99US-0144884.
PR	28-APR-1999;	99US-0131449.	PR	21-JUL-1999;	99US-0148114.
PR	30-APR-1999;	99US-0132048.	PR	21-JUL-1999;	99US-0145086.
PR	30-APR-1999;	99US-0132407.	PR	21-JUL-1999;	99US-0145088.
PR	04-MAY-1999;	99US-0132484.	PR	22-JUL-1999;	99US-0145085.
PR	05-MAY-1999;	99US-0132485.	PR	22-JUL-1999;	99US-0145087.
PR	06-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-0145089.
PR	07-MAY-1999;	99US-0132487.	PR	23-JUL-1999;	99US-0145192.
PR	07-MAY-1999;	99US-0132863.	PR	23-JUL-1999;	99US-0145145.
PR	11-MAY-1999;	99US-0134256.	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134218.	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134219.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134221.	PR	27-JUL-1999;	99US-0145913.
PR	14-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145918.
PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145919.
PR	19-MAY-1999;	99US-0134941.	PR	28-JUL-1999;	99US-0145951.
PR	20-MAY-1999;	99US-0135124.	PR	02-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146388.
PR	21-MAY-1999;	99US-0135629.	PR	02-AUG-1999;	99US-0146389.
PR	24-MAY-1999;	99US-0136021.	PR	03-AUG-1999;	99US-0147038.
PR	25-MAY-1999;	99US-0136392.	PR	04-AUG-1999;	99US-0147204.
PR	27-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147302.
PR	28-MAY-1999;	99US-0137222.	PR	05-AUG-1999;	99US-0147192.
PR	01-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	03-JUN-1999;	99US-0137528.	PR	06-AUG-1999;	99US-0147303.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147416.
PR	07-JUN-1999;	99US-0137724.	PR	09-AUG-1999;	99US-0147493.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138540.	PR	10-AUG-1999;	99US-0148171.
PR	10-JUN-1999;	99US-0138847.	PR	11-AUG-1999;	99US-0148319.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148565.
PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148684.
PR	17-JUN-1999;	99US-0139452.	PR	13-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	23-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139462.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	21-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140595.	PR	07-SEP-1999;	99US-0152363.
PR	24-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	28-JUN-1999;	99US-0140823.	PR	13-SEP-1999;	99US-0153758.
PR	29-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0154018.
PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.

Db 162 tggrrgggykgsdfvvgihggilifballnslpsivlsfigqlaalwollgvlvmlili 221
QY 209 SFIAILARSNPKANSOQWTAWSNYTG---WSDGVCIFILGLSTSCFPGICLDAAMHAAEE 265
Db 222 plv-----steratktkfntfndnglgtisyayifvlglmsqylitgydasahmtec 276
QY 266 CTDAAITVPKAVNSAIIIGECTAFPTIAYLYGTDLDSILS---SAGYIPFETWTQSL 321
Db 277 tvdadkngpgrilsaigslilfgwgylilgisyavtdilpsilsetnnggyaiaefylaf 336
QY 322 RSL--SFATVLSGGI--VMAFALNAVQETASRLTWSFARDNGLVSTHLRIHPRWQV 377
Db 337 knrfsgstggivclgavavafcgmsvtsnsrmayafsrddgampmsplwhkvnar-ev 395
QY 378 PVWSLFATW--GILATCGCI-FLGSAFNAVNSAVVLOQLSFLPIALLLLYQKRDPKF 434
Db 396 pl--navwlsalifcmaltsgsivafqamvsiatigilyayaipilrvltarn-tf 451
QY 435 LPSTRAFLPRGIGFLNVAVLAVFTSVTVTFPSFPLTVPTAASTMYNTSAIIGVALALGV 494
Db 452 vpg--prslgk-ygmvgvawivwvtisvlfpvaypitaetlnytpvavaglvaitl 508
QY 495 LNWVYHARKHYQCP 508
Db 509 sywlfsarhwftgp 522
RESULT 9
AAG31988
ID AAG31988 standard; Protein; 438 AA.
XX AC AAG31988;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 38508.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EPI033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 18-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 99US-0134219.
PR 99US-0134221.
PR 99US-0134370.
PR 14-MAY-1999; 99US-0134768.
PR 18-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 28-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151063.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159364.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.

PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.
 Query Match 14.3%; Score 400; DB 21; Length 438;
 Best Local Similarity 27.4%; Pred. No. 1.le-32;
 Matches 122; Conservative 93; Mismatches 191; Indels 40; Gaps 16;
 Qy 89 ICIAFSIAELTSVYPTAGQYHEASILAPKSINRSISYV--CGLVSLLSWIAIGSSVTMI 146
 Db 1 mcvglsmaeicssyptsqggylywsamlagrpwaplaswmtgckvvclqW-avtasvdfs 59
 Qy 147 PAQOIPALI-----AAYSHTYSQDSWHVFLIYEGVALVLLFNLFALKRNPWYHEIG--- 198
 Db 60 laqliqvllstggrngggykgsdfvvgihggilfihallnslpsvlsfigqialw 119
 Qy 199 --FGLTIALFVIFIALARSNPKAPNSQWTAWSNWTG---WSDGVCFILGLSTSCFMP 253
 Db 120 nllgvlvmlilply-----sterattkfvftntdnglgtisyayifvlgllmsqyti 174
 Qy 254 IGLDAAMHLAECTDAARTVPKAVVSAILIGFCTAFPTTIAYLYGITDLSILS----SA 309
 Db 175 tgydasahmtteetvdadkngprgiisaigilsilfgwgyligisyavtdlpsllsetnng 234
 Qy 310 GYIFETWTQSLRSL--SPATVLSGGI--VMAFFALNAVOETASRLTWSFARDNGLVPS 365
 Db 235 gyalaeifylafknrfsgtggivclgvavavffcgmsvtsnrmayafsrddgamps 294
 Qy 366 TLERIHPRMQVPVMSLFATW--GILATCGCI-FLGSTAFNALVNSAVVLOQLSLPI 422
 Db 295 plwhkvnsr-evpi---navwlsalisfcmaltsglsivafqamvsiatiglyayaipi 350
 Qy 423 ALLLYQRDKPFLPSTRAFLVLPRIIGLVNVLAVVFTSVTFVFPFLVPTTAASMTNNT 482
 Db 351 ilrvtlarn-tfvpqg--pfslgk-ygmvgvavvlwvvtisvlfsipvaypitaetinyt 406
 Qy 483 SAIIGVALALGVLMNVVHARKHYQGP 508
 Db 407 pvavaglvaitlsyflisarhwftgp 432

RESULT 10
 AAG49886
 ID AAG49886 standard; Protein; 437 AA.
 XX AAG49886;
 AC AAG49886;
 DT 18-OCT-2000 (first entry)
 XX Arabidopsis thaliana protein fragment SEQ ID NO: 63156.
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 63156.
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX Arabidopsis thaliana.
 OS Arabidopsis thaliana.
 XX EP1033405-A2.
 XX 06-SEP-2000.
 XX 25-FEB-2000; 2000EP-0301439.
 XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.


```
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 14.8%; Score 398.5; DB 21; Length 437;
Best local similarity 26.8%; Pred. No. 1.6e-32;
Matches 119; Conservative 95; Mismatches 193; Indels 37; Gaps 14;

Oy 89 ICIAFSAEITSVYPTAGGQYHFPASIIAPKSIINRSISYVCGLVSLLSWIAIGSSVYTMIPA 148
Db 1 mcvglsmaeicssyptsqgllywsamlagrpwaplasmtgtnfnvgqwavtasvdfsla 60

Oy 149 QOIPALI-----AAYSHYSODSMHWFLIYEGVALVLLFNLFALKRNPWVHEIG----- 198
Db 61 qlivqivllstgrnggykgsdfvvlghgflfihallnslpislvsiflglaalwnl 120

Oy 199 FGLTIALFVLSFTAILARSPKAPNSQVMTAWSNYTG---WSDGVCFILGLSTSCFMFTG 255
Db 121 lgvilmlilplv-----sterattkfvfnfntdnglgtisyayifvlgllmsqytlg 175

Oy 256 LDAAMHAEECTDAARTVPRKAVVSAILIIGFCTAFPTTIAVLXGITDLSILS-----SAGY 311
Db 176 ydasahmteetvdadkngprgiisaigisilfgwgyilglsyavtdipallsetnnsgy 235

Oy 312 IPFETMTQSLRSL--SPATVLSGGI--VMAFFALNAVOETASRLTWSFARONGLVFSTH 367
Db 236 aiaaiefylaknrrfsgtgdvclgvavavfvcgmsvtsnrmayafrdgdampmspl 295

Oy 368 LERHPRMQVPSLFW--GILATGCCI--FLGSTAFNALNSAVVLOQLSFLPIAL 424
Db 296 whkvnsr-evpi---navwisaliscfmaltsigsivafqamvsiatiglyayaipil 351

Oy 425 LLYOKRQPKLPSPRAVPLPRGIFGLVNVLAVFTSVTVTFEFPLTVPTAASTWNYS 484
Db 352 rvtlarn-tfvpg--pfslqk-ygmvmvgwvavivmtvisvlflpvaypitaetlnytpv 407

Oy 485 IIGVALGLVNLVWVHARKHYQGP 508
Db 408 avaglvaitlsywlfshwftgp 431
```

RESULT 11
AAG49911
ID AAG49911 standard; Protein; 439 AA.
XX

```
AC AAG49911;
XX 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 63191.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 23-APR-1999; 99US-0130891.
XX PR 28-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 30-APR-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 06-MAY-1999; 99US-0132487.
XX PR 07-MAY-1999; 99US-0132863.
XX PR 11-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 14-MAY-1999; 99US-0134370.
XX PR 18-MAY-1999; 99US-0134768.
XX PR 19-MAY-1999; 99US-0134941.
XX PR 20-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
XX PR 25-MAY-1999; 99US-0136021.
XX PR 27-MAY-1999; 99US-0136392.
XX PR 28-MAY-1999; 99US-0136782.
XX PR 01-JUN-1999; 99US-0137222.
XX PR 03-JUN-1999; 99US-0137528.
XX PR 04-JUN-1999; 99US-0137502.
XX PR 07-JUN-1999; 99US-0137724.
XX PR 08-JUN-1999; 99US-0138094.
XX PR 10-JUN-1999; 99US-0138540.
XX PR 10-JUN-1999; 99US-0138847.
XX PR 14-JUN-1999; 99US-0139119.
XX PR 16-JUN-1999; 99US-0139452.
XX PR 16-JUN-1999; 99US-0139453.
XX PR 17-JUN-1999; 99US-0139452.
XX PR 18-JUN-1999; 99US-0139454.
XX PR 18-JUN-1999; 99US-0139455.
XX PR 18-JUN-1999; 99US-0139456.
XX PR 18-JUN-1999; 99US-0139457.
XX PR 18-JUN-1999; 99US-0139458.
XX PR 18-JUN-1999; 99US-0139459.
XX PR 18-JUN-1999; 99US-0139460.
XX PR 18-JUN-1999; 99US-0139461.
XX PR 18-JUN-1999; 99US-0139462.
```


PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 25-MAY-1999; 99US-0136392.
PR 27-MAY-1999; 99US-0136782.
PR 28-MAY-1999; 99US-0137222.
PR 01-JUN-1999; 99US-0137528.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137724.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140595.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0130891.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0153263.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2002, 20:50:14 ; Search time 5001.19 Seconds
(without alignments)
4258.625 Million cell updates/sec

Title: US-09-882-694A-7
Perfect score: 1578
Sequence: 1 atggactccagaccacgaatgg.....aattccaagttgggccatga 1578

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
c 1	55	3.5	1010	12	CNS06EC5	AL394923 T7 end of
c 2	54	3.4	521	12	AZ928850	AZ928850 479.dif18
c 3	51.2	3.2	939	12	CNS07DIC	AL440506 T7 end of
c 4	49.8	3.2	686	10	BT954357	BT954357 HVSMEM001
c 5	49.6	3.1	634	9	AW442992	AW442992 EST307922
c 6	49	3.1	948	12	CNS06DG4	AL393770 T7 end of
c 7	48.8	3.1	525	12	AQ397336	AQ397336 mgxb0012D
c 8	46.2	2.9	644	12	AQ161795	AQ161795 mgxb00091
c 9	43	2.7	1028	12	CNS06XLB	AL419877 T3 end of
c 10	42.6	2.7	511	10	BG241224	BG241224 FVL_39_F1
c 11	42.6	2.7	546	10	BF586115	BF586115 FVL_26_H0
c 12	42.6	2.7	997	12	CNS005TE	AL060767 Drosophil
c 13	42.4	2.7	375	9	AY784308	AY784308 d4c12a1.f
c 14	42.2	2.7	679	10	BG810178	BG810178 mgct0003xe
c 15	41.6	2.6	385	9	AU031682	AU031682 AU031682
c 16	41.6	2.6	675	10	BT954997	BT954997 HVSMEM002
c 17	41	2.6	601	10	BG463370	BG463370 EM1_48_F1

18	41	2.6	663	10	BI955572	BI955572 HVSMEM002
c 19	40.6	2.6	641	12	CNS07NX	AL432931 T7 end of
c 20	40.2	2.5	1101	12	CNS012S8	AL101954 Drosophil
c 21	39.8	2.5	1273	12	AQ896271	AQ896271 HS_3176_B
c 22	39.6	2.5	969	12	CNS02S69	AL211482 Tetraodon
c 23	39.2	2.5	383	10	BE523754	BE523754 M41C3STM
c 24	39.2	2.5	443	9	AA788013	AA788013 r4a12a1.r
c 25	39.2	2.5	450	9	AV548940	AV548940 AV548940
c 26	37.4	2.4	399	9	AW872351	AW872351 hm28c11.x
c 27	37.4	2.4	639	9	AL648272	AL648272 AL648272
c 28	37.2	2.4	298	10	BE936629	BE936629 QV4-OT003
c 29	37	2.3	913	12	CNS079UV	AL435773 T3 end of
c 30	36.8	2.3	517	12	AZ930479	AZ930479 474.dhz54
c 31	36.8	2.3	748	10	BT970431	BT970431 GM830010B
c 32	36.6	2.3	256	9	BB449103	BB449103 BB449103
c 33	36.6	2.3	490	10	BT773560	BT773560 rp03b12.y
c 34	36.6	2.3	586	10	BF255233	BF255233 HVSMEM000
c 35	36.6	2.3	956	12	CNS01KTF	AL148772 Anopheles
c 36	36.4	2.3	548	9	AI660809	AI660809 we69a11.x
c 37	36.4	2.3	548	9	AA477575	AA477575 zu44b02.r
c 38	36.4	2.3	793	12	BH111627	BH111627 RPCI-24-2
c 39	36.4	2.3	1650	12	AG150948	AG150948 Pan trogl
c 40	36.2	2.3	328	10	D23882	D23882 R1CR0503A.R
c 41	36.2	2.3	469	9	AA655635	AA655635 vs46h11.r
c 42	36.2	2.3	867	12	A2126336	A2126336 OSJNB007
c 43	36	2.3	394	10	BI776502	BI776502 EBPI105_SQ
c 44	36	2.3	408	9	AW397789	AW397789 sg68d06.y
c 45	36	2.3	472	9	AL378484	AL378484 MLBB38E05

ALIGNMENTS

CNS06EC5 1010 bp DNA linear GSS 17-JUN-2001
T7 end of clone AR0AA020D08 of library AR0AA from strain CBS 732 of
Zygosaccharomyces rouxii, genomic survey sequence.

AL394923 1 GI:12146190
GSS.
Zygosaccharomyces rouxii.
Zygosaccharomyces rouxii
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Zygosaccharomycetes.
1 (bases 1 to 1010)
de Montigny, J., Straub, M., Potier, S., Tekala, F., Dujon, B.,
Wincker, P., Artiguenave, F. and Souciet, J.
Genomic exploration of the hemiascomycetous yeasts: 8.
Zygosaccharomyces rouxii
FEBS Lett. 487 (1), 52-55 (2000)
20584718
2 (bases 1 to 1010)
Souciet, J., Aigle, M., Artiguenave, F., Blandin, G.,
Boletoin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,
de Montigny, J., Dujon, B., Durand, P., Lepingle, A., Liorente, B.,
Maupertuy, A., Neveu, G., Ozier-Kalogirou, O., Potier, S.,
Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
Wincker, P. and Weissbach, J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
20584711
3 (bases 1 to 1010)
Direct Submission
Genoscope.
Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces

barley genomics. Barley Genetics Newsletter 31:29-30.
(<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>.)"

152 a 170 c 186 g 178 t

BASE COUNT


```

QY 1461 cgttgcaactgctgctgtgctgaactggtgctgcatgcccagagcattatcagg 1520
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 296 GGCCCTCTTCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 355
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1521 acccaacttgagcttgacgacgggtgc 1549
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 356 CCCCCTCACCAATCTCGACGGATGAATG 384
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
CN5005TE
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR12K22 of RPCL-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster
fruit fly.
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 997)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segre@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org/TheBDGP/Drosophila
melanogaster BAC library was prepared by Kazuoto Osoegawa and
Aaron Mamoser in Peter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCL-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
PI and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
FEATURES
source
1..997
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCL-98"
/clone="BACR12K22"
/note="end : TET3"
BASE COUNT 89 a 99 c 13 g 258 t 538 others
ORIGIN

Query Match 2.7%; Score 42.6; DB 12; Length 997;
Best Local Similarity 16.5%; Pred. No. 0.7;
Matches 82; Conservative 175; Mismatches 240; Indels 0; Gaps 0;

QY 910 tctattcaagctcgcgcgtatattccattcgagacaaatgacgagcttcggtcg 969
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 484 TTTTYYTCTTYYTCTTYYTCTTYYTCTTYYTCTTYYTCTTYYTCTTYYTCTT 543
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 970 ctcagtttgcaaggtctctatgtgagcggtatgagcctcttcgcccctaac 1029
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 544 TTTTYYTCTTYYTCTTYYTCTTYYTCTTYYTCTTYYTCTTYYTCTTYYTCTT 603
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1030 gctgtacaagagactgctctcactcactgagcttgcgcgcgagacaaatggctgta 1089
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 604 YCVYYCYCTTYYTCTTYYTCTTYYTCTTYYTCTTYYTCTTYYTCTTYYTCTT 663
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1090 ttctccactcatctcgaacgcatcattcccccgtcgcgaacttctgttctctatc 1149
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Query Match 2.1% Score 32.6; DB 6; Length 1734;
Best Local Similarity 55.9%; Pred. No. 3;
Matches 62; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

PRIOR APPLICATION NUMBER: 60/103711
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/104257
PRIOR FILING DATE: 1998-10-14
PRIOR APPLICATION NUMBER: 60/104987
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105000
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105002
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105104
PRIOR FILING DATE: 1998-10-21
PRIOR APPLICATION NUMBER: 60/105169
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105266
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105693
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105694
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105807
PRIOR FILING DATE: 1998-10-27
PRIOR APPLICATION NUMBER: 60/105881
PRIOR FILING DATE: 1998-10-27
PRIOR APPLICATION NUMBER: 60/105882
PRIOR FILING DATE: 1998-10-27
PRIOR APPLICATION NUMBER: 60/106023
PRIOR FILING DATE: 1998-10-28
PRIOR APPLICATION NUMBER: 60/106029

US-10-006-768A-51/c
Sequence 51, Application US/10006768A
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2830PIC10
CURRENT FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: 60/1006768A
PRIOR FILING DATE: 2002-03-05
Number of SEQ ID NOS: 477
Prior Application removed - See File Wrapper or Palm
SEQ ID NO 51
LENGTH: 1734
TYPE: DNA
ORGANISM: Homo sapiens
US-10-006-768A-51

Query Match 2.1% Score 32.6; DB 6; Length 1734;
Best Local Similarity 55.9%; Pred. No. 3;
Matches 62; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

PRIOR APPLICATION NUMBER: 60/103711
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/104257
PRIOR FILING DATE: 1998-10-14
PRIOR APPLICATION NUMBER: 60/104987
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105000
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105002
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105104
PRIOR FILING DATE: 1998-10-21
PRIOR APPLICATION NUMBER: 60/105169
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105266
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105693
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105694
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105807
PRIOR FILING DATE: 1998-10-27
PRIOR APPLICATION NUMBER: 60/105881
PRIOR FILING DATE: 1998-10-27
PRIOR APPLICATION NUMBER: 60/105882
PRIOR FILING DATE: 1998-10-27
PRIOR APPLICATION NUMBER: 60/106023
PRIOR FILING DATE: 1998-10-28
PRIOR APPLICATION NUMBER: 60/106029

US-10-006-768A-51/c
Sequence 51, Application US/10006768A
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2830PIC10
CURRENT FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: 60/1006768A
PRIOR FILING DATE: 2002-03-05
Number of SEQ ID NOS: 477
Prior Application removed - See File Wrapper or Palm
SEQ ID NO 51
LENGTH: 1734
TYPE: DNA
ORGANISM: Homo sapiens
US-10-006-768A-51

Query Match 2.1% Score 32.6; DB 6; Length 1734;
Best Local Similarity 55.9%; Pred. No. 3;
Matches 62; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

PRIOR APPLICATION NUMBER: 60/103711
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/104257
PRIOR FILING DATE: 1998-10-14
PRIOR APPLICATION NUMBER: 60/104987
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105000
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105002
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105104
PRIOR FILING DATE: 1998-10-21
PRIOR APPLICATION NUMBER: 60/105169
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105266
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105693
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105694
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105807
PRIOR FILING DATE: 1998-10-27
PRIOR APPLICATION NUMBER: 60/105881
PRIOR FILING DATE: 1998-10-27
PRIOR APPLICATION NUMBER: 60/105882
PRIOR FILING DATE: 1998-10-27
PRIOR APPLICATION NUMBER: 60/106023
PRIOR FILING DATE: 1998-10-28
PRIOR APPLICATION NUMBER: 60/106029

US-10-017-527A-51/c
Sequence 51, Application US/10017527A
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2830PIC63
CURRENT FILING DATE: 2001-12-13
PRIOR APPLICATION NUMBER: 60/098716
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098723
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098749
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098750
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098803
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098821
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098843
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/099536
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099596
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099598
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099602
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099642
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099741
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099754
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099763
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099792
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099808
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099812

QY 698 atactgctgtgcacagcgcttctcttcattccctggccttttcgacatctctctcatgt 757
 ||| | ||| ||||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1229 ATCCAGATTCCTCGCTCCGCGGGCTTCAATCCCTGCCTTTTCACACCCGGGTATTGT 1170
 ||| | ||| ||||| ||| ||| ||| ||| ||| ||| ||| |||
QY 758 tcaattgcttgacagcagaatgatcatctgctgaagaatagcacagtgtgtg 808
 |||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1169 CCATTTCCTCCGCGCTCACCACCGTGCTTGC CGAGGAGGAGCGGTGCTG 1119
 |||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
US-09-540-209B-4701
; Sequence 4701, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BANA
; FILE REFERENCE: FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09-540-209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 4701
; LENGTH: 705
; TYPE: DNA
; ORGANISM: B.fragilis
US-09-540-209B-4701

Query Match	2.0%;	Score 31.2;	DB 5;	Length 705;
Best Local Similarity	47.9%;	Pred. No. 5.5;		
Matches	90;	Conservative	0;	Mismatches 98;
				Indels 0;
				Gaps

QY 1364 cggtagcttacacgtccogtcacacaactgtgttttttcagctcttcacactgacgctgctaagg 1423
 |||| |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 365 cgdgtactattcacggcgtggtctctcttatcttctgccctcccagggacaccattcacca 424
 |||| |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1424 ccgcgtcaaccatgaattacacaaagtcgattatagggttgacgttgtctgtgtgtgtgtct 1483
 |||| |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 425 ccgtggcatgatgatgcgcgttgctgctgatgcattcccgtagaacgttggtgtacctac 484
 |||| |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1484 tgaactgggtgctgctgcatgccaggaagcattatcaggagccccacttgagcttgacggac 1543
 |||| |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 485 tgctttacctctctccggagaaggtgtcggtggaaactcttcttctcacttaacgcgat 544
 |||| |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1544 ggggtcgtc 1551
 |||| ||| |||
Db 545 tggtagcc 552

RESULT 10
US-09-866-557A-3/C
; Sequence 3, Application US/09866557A
; GENERAL INFORMATION:
; APPLICANT: Hammond, S.
; APPLICANT: Hannon, G.
; APPLICANT: Beach, D.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR RNA INTERFERENCE
; FILE REFERENCE: GNCA-P02-007
; CURRENT APPLICATION NUMBER: US/09/866,557A
; CURRENT FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: 60/189,739
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/243,097
; PRIOR FILING DATE: 2000-10-24
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: patentin version 3.1
; SEQ ID NO 3
; LENGTH: 6750
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(6750)
; OTHER INFORMATION:


```
RESULT 15
US-10-003-671A-14
; Sequence 14, Application US/10003671A
; GENERAL INFORMATION:
; APPLICANT: MIZE, ET AL
; TITLE OF INVENTION: NOVEL INTERLEUKIN-1 HY2 MATERIALS AND METHODS
; FILE REFERENCE: 28110/36858A
; CURRENT APPLICATION NUMBER: US/10/003,671A
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/245,346
; PRIOR FILING DATE: 2000-11-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 1366
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-003-671A-14

Query Match      1.9%; Score 30.4; DB 6; Length 1366;
Best Local Similarity 55.8%; Pred. NO. 13;
Matches 58; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 1152 gacctgggggaattctagccacatgcggatgtatatctctaggttctagcacagctttcaa 1211
      |||||
Db 284 ggcttg999gaaccogtgcagccogtggctgagtggtttctaaagcccaagcaagctgcctc 343
      |||||

QY 1212 tgccttggtcaattccggcgttggactccagcaactctctcc 1255
      |||||
Db 344 tggcttcaccagcctctcttcttaactgcctctctctctcc 387
      |||||
```

Search completed: April 27, 2002, 01:37:27
Job time: 17180 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 01:33:51 ; Search time 8172.36 Seconds

(without alignments)
4177.480 Million cell updates/sec

Title: US-09-882-694A-7

Perfect score: 1578

Sequence: 1 agggactcagaccaaagtgaattcaagttgggccatga 1578

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2179536 seqs, 10817449327 residues

Total number of hits satisfying chosen parameters: 43959072

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents.NA.Main.*

- 1: /cgn2_6/ptodata/2/pna/US085_COMB.seq.*
- 2: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
- 3: /cgn2_6/ptodata/2/pna/US087_COMB.seq.*
- 4: /cgn2_6/ptodata/2/pna/US088_COMB.seq.*
- 5: /cgn2_6/ptodata/2/pna/US089_COMB.seq.*
- 6: /cgn2_6/ptodata/2/pna/US090_COMB.seq.*
- 7: /cgn2_6/ptodata/2/pna/US091_COMB.seq.*
- 8: /cgn2_6/ptodata/2/pna/US092_COMB.seq.*
- 9: /cgn2_6/ptodata/2/pna/US093_COMB.seq.*
- 10: /cgn2_6/ptodata/2/pna/US094_COMB.seq.*
- 11: /cgn2_6/ptodata/2/pna/US095A_COMB.seq.*
- 12: /cgn2_6/ptodata/2/pna/US095B_COMB.seq.*
- 13: /cgn2_6/ptodata/2/pna/US095C_COMB.seq.*
- 14: /cgn2_6/ptodata/2/pna/US095D_COMB.seq.*
- 15: /cgn2_6/ptodata/2/pna/US095E_COMB.seq.*
- 16: /cgn2_6/ptodata/2/pna/US095F_COMB.seq.*
- 17: /cgn2_6/ptodata/2/pna/US095G_COMB.seq.*
- 18: /cgn2_6/ptodata/2/pna/US095H_COMB.seq.*
- 19: /cgn2_6/ptodata/2/pna/US095I_COMB.seq.*
- 20: /cgn2_6/ptodata/2/pna/US095J_COMB.seq.*
- 21: /cgn2_6/ptodata/2/pna/US095K_COMB.seq.*
- 22: /cgn2_6/ptodata/2/pna/US095L_COMB.seq.*
- 23: /cgn2_6/ptodata/2/pna/US095M_COMB.seq.*
- 24: /cgn2_6/ptodata/2/pna/US095N_COMB.seq.*
- 25: /cgn2_6/ptodata/2/pna/US095O_COMB.seq.*
- 26: /cgn2_6/ptodata/2/pna/US095P_COMB.seq.*
- 27: /cgn2_6/ptodata/2/pna/US095Q_COMB.seq.*
- 28: /cgn2_6/ptodata/2/pna/US095R_COMB.seq.*
- 29: /cgn2_6/ptodata/2/pna/US095S_COMB.seq.*
- 30: /cgn2_6/ptodata/2/pna/US095T_COMB.seq.*
- 31: /cgn2_6/ptodata/2/pna/US095U_COMB.seq.*
- 32: /cgn2_6/ptodata/2/pna/US095V_COMB.seq.*
- 33: /cgn2_6/ptodata/2/pna/US095W_COMB.seq.*
- 34: /cgn2_6/ptodata/2/pna/US095X_COMB.seq.*
- 35: /cgn2_6/ptodata/2/pna/US095Y_COMB.seq.*
- 36: /cgn2_6/ptodata/2/pna/US095Z_COMB.seq.*
- 37: /cgn2_6/ptodata/2/pna/US096_COMB.seq.*
- 38: /cgn2_6/ptodata/2/pna/US096A_COMB.seq.*
- 39: /cgn2_6/ptodata/2/pna/US096B_COMB.seq.*
- 40: /cgn2_6/ptodata/2/pna/US096C_COMB.seq.*
- 41: /cgn2_6/ptodata/2/pna/US096D_COMB.seq.*
- 42: /cgn2_6/ptodata/2/pna/US096E_COMB.seq.*
- 43: /cgn2_6/ptodata/2/pna/US096F_COMB.seq.*

- 44: /cgn2_6/ptodata/2/pna/US6005_COMB.seq.*
- 45: /cgn2_6/ptodata/2/pna/US6006_COMB.seq.*
- 46: /cgn2_6/ptodata/2/pna/US6007_COMB.seq.*
- 47: /cgn2_6/ptodata/2/pna/US6008_COMB.seq.*
- 48: /cgn2_6/ptodata/2/pna/US6009_COMB.seq.*
- 49: /cgn2_6/ptodata/2/pna/US6010_COMB.seq.*
- 50: /cgn2_6/ptodata/2/pna/US6011_COMB.seq.*
- 51: /cgn2_6/ptodata/2/pna/US6012_COMB.seq.*
- 52: /cgn2_6/ptodata/2/pna/US6013_COMB.seq.*
- 53: /cgn2_6/ptodata/2/pna/US6014_COMB.seq.*
- 54: /cgn2_6/ptodata/2/pna/US6015_COMB.seq.*
- 55: /cgn2_6/ptodata/2/pna/US6016_COMB.seq.*
- 56: /cgn2_6/ptodata/2/pna/US6017_COMB.seq.*
- 57: /cgn2_6/ptodata/2/pna/US6018_COMB.seq.*
- 58: /cgn2_6/ptodata/2/pna/US6019_COMB.seq.*
- 59: /cgn2_6/ptodata/2/pna/US6020_COMB.seq.*
- 60: /cgn2_6/ptodata/2/pna/US6021_COMB.seq.*
- 61: /cgn2_6/ptodata/2/pna/US6022_COMB.seq.*
- 62: /cgn2_6/ptodata/2/pna/US6023_COMB.seq.*
- 63: /cgn2_6/ptodata/2/pna/US6024_COMB.seq.*
- 64: /cgn2_6/ptodata/2/pna/US6025_COMB.seq.*
- 65: /cgn2_6/ptodata/2/pna/US6026_COMB.seq.*
- 66: /cgn2_6/ptodata/2/pna/US6027_COMB.seq.*
- 67: /cgn2_6/ptodata/2/pna/US6028_COMB.seq.*
- 68: /cgn2_6/ptodata/2/pna/US6029_COMB.seq.*
- 69: /cgn2_6/ptodata/2/pna/US6030_COMB.seq.*
- 70: /cgn2_6/ptodata/2/pna/US6031_COMB.seq.*
- 71: /cgn2_6/ptodata/2/pna/US6032_COMB.seq.*
- 72: /cgn2_6/ptodata/2/pna/US6033_COMB.seq.*
- 73: /cgn2_6/ptodata/2/pna/US6034_COMB.seq.*
- 74: /cgn2_6/ptodata/2/pna/US6035_COMB.seq.*
- 75: /cgn2_6/ptodata/2/pna/US6036_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1578	100.0	1578	US-09-351-224E-7	Sequence 7, Appli
2	1578	100.0	1578	US-09-677-488A-7	Sequence 7, Appli
3	1578	100.0	1578	US-09-677-682A-7	Sequence 7, Appli
4	1578	100.0	1578	US-09-677-682B-7	Sequence 7, Appli
5	1578	100.0	1578	US-09-882-694-7	Sequence 7, Appli
6	1578	100.0	1578	US-09-882-694A-7	Sequence 7, Appli
7	1578	100.0	1764	US-09-351-224E-6	Sequence 6, Appli
8	1578	100.0	1764	US-09-677-488A-6	Sequence 6, Appli
9	1578	100.0	1764	US-09-677-682A-6	Sequence 6, Appli
10	1578	100.0	1764	US-09-677-682B-6	Sequence 6, Appli
11	1578	100.0	1764	US-09-882-694A-6	Sequence 6, Appli
12	1555.6	98.6	1578	US-09-351-224-7	Sequence 7, Appli
13	1555.6	98.6	1578	US-09-351-823-7	Sequence 7, Appli
14	1555.6	98.6	1578	US-09-677-488-7	Sequence 7, Appli
15	1555.6	98.6	1578	US-09-677-682-7	Sequence 7, Appli
16	1555.6	98.6	1764	US-09-351-224-6	Sequence 6, Appli
17	1555.6	98.6	1764	US-09-677-488-6	Sequence 6, Appli
18	1555.6	98.6	1764	US-09-677-682-6	Sequence 6, Appli
19	1555.6	98.6	1764	US-09-882-694-6	Sequence 6, Appli
20	1555.6	98.6	1764	US-09-882-694A-6	Sequence 6, Appli
21	106.8	6.8	609	US-09-417-507-2341	Sequence 2341, Ap
22	91.8	5.8	1632	US-09-404-520-15131	Sequence 15131, A
23	90.2	5.7	837	US-06-138-103-7348	Sequence 7348, Ap
24	89.4	5.7	483	US-09-417-507-2340	Sequence 2340, Ap
25	86.2	5.5	1466	US-09-404-520-19952	Sequence 19952, A
26	86.2	5.5	1504	US-60-138-103-5913	Sequence 5913, Ap
27	86.2	5.5	1734	US-09-404-520-1908	Sequence 1908, Ap
28	84.2	5.3	807	US-08-675-784A-72	Sequence 72, Appli
29	81.2	5.1	864	US-09-417-507-15475	Sequence 15475, A
30	79.2	5.0	1980	US-60-082-300-10911	Sequence 10911, A
31	78.2	5.0	1981	US-09-404-520-20192	Sequence 20192, A

32 78.2 5.0 2734 18 US-09-404-520-20302 Sequence 20302, A
C 34 78.2 5.0 5837 52 US-60-138-103-7923 Sequence 7923, Ap
C 34 78.2 5.0 7194 18 US-09-404-520-3633 Sequence 3633, Ap
35 76.6 4.9 549 18 US-09-417-507-2343 Sequence 2343, Ap
36 75.8 4.8 1194 18 US-09-404-520-19708 Sequence 19708, A
37 75.8 4.8 1420 18 US-09-404-520-19922 Sequence 19922, A
C 38 75.8 4.8 1517 52 US-60-138-103-3579 Sequence 3579, Ap
C 39 75.8 4.8 5225 18 US-09-404-520-4899 Sequence 4899, Ap
40 73.2 4.6 1059 18 US-09-417-507-17478 Sequence 17478, A
41 71 4.5 1236 18 US-09-417-507-11427 Sequence 11427, A
C 42 71 4.5 1293 18 US-09-417-507-11429 Sequence 11429, A
C 43 69.8 4.4 585 18 US-09-417-507-17798 Sequence 17798, A
C 44 68.2 4.3 696 47 US-60-082-300-5633 Sequence 5633, Ap
45 68.2 4.3 2602 18 US-09-404-520-20294 Sequence 20294, A

ALIGNMENTS

RESULT 1
US-09-351-224E-7
: Sequence 7, Application US/09351224E
: GENERAL INFORMATION:
: APPLICANT: DuVick, Jon
: APPLICANT: Maddox, Joyce
: APPLICANT: Gilliam, Jacob
: APPLICANT: Folkerts, Otto
: APPLICANT: Crasta, Oswald R.
: TITLE OF INVENTION: Compositions and Methods for Fumonisin
: FILE REFERENCE: 5718-111
: CURRENT APPLICATION NUMBER: US/09/351.224E
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 7
: LENGTH: 1578
: TYPE: DNA
: ORGANISM: Exophiala spinifera
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (0)..(0)
: OTHER INFORMATION: permealase, fully spliced cDNA
US-09-351-224E-7

Query Match 100.0%; Score 1578; DB 17; Length 1578;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atggactccagaccagtggtatcgagcgagaaagcgagagcgagacgaagaac 60
Db 1 atggactccagaccagtggtatcgagcgagaaagcgagagcgagacgaagaac 60
QY 61 acagagacggcgagcgagtggtggtcgagtcctcgagtccttccttggaaga 120
Db 61 acagagacggcgagcgagtggtggtcgagtcctcgagtccttccttggaaga 120
QY 121 caatttggcaccatcacatcgatcgcttggccttggcttggatggcaggtggctgt 180
Db 121 caatttggcaccatcacatcgatcgcttggccttggcttggatggcaggtggctgt 180
QY 181 atctcaggcagcttcacatcgctcctactagcgggggggggcctcctcttcaggc 240
Db 181 atctcaggcagcttcacatcgctcctactagcgggggggggcctcctcttcaggc 240
QY 241 atctcgaatcagtaactcgttacatcgatcgcttcttcattgacgaactgaccagc 300
Db 241 atctcgaatcagtaactcgttacatcgatcgcttcttcattgacgaactgaccagc 300
QY 301 gtctaccgactgcccggggcgaatacatcttctgctcgatcctggcaccacaaatc 360
Db 301 gtctaccgactgcccggggcgaatacatcttctgctcgatcctggcaccacaaatc 360

QY 361 aatcgagacattcacaacgtgtcgaggaactcggtgctgtgtcttctcatgcatcgctatcgga 420
Db 361 aatcgagacattcacaacgtgtcgaggaactcggtgctgtgtcttctcatgcatcgctatcgga 420
QY 421 agctcagtgacacatgatacctgctcaacagatccccggcgctgatacagcgccctatagtcaac 480
Db 421 agctcagtgacacatgatacctgctcaacagatccccggcgctgatacagcgccctatagtcaac 480
QY 481 acatactcccagagatcggtgcatgtcttctcctcaatctacagagggagtcgctggtggtg 540
Db 481 acatactcccagagatcggtgcatgtcttctcctcaatctacagagggagtcgctggtggtg 540
QY 541 ctctgttcaactgtttgcccctgaaagaaacccctgggttctcagaaatcggaatctcggc 600
Db 541 ctctgttcaactgtttgcccctgaaagaaacccctgggttctcagaaatcggaatctcggc 600
QY 601 ctacgatacgtctctctgtgatctcttcttctgcacattctctagcggcggtcccaaccccaag 660
Db 601 ctacgatacgtctctctgtgatctcttcttctgcacattctctagcggcggtcccaaccccaag 660
QY 661 gctccaaactcacagatgtaggactgcttgagcaactatactactggtgctgctccagagggctc 720
Db 661 gctccaaactcacagatgtaggactgcttgagcaactatactactggtgctgctccagagggctc 720
QY 721 tgcttcaactcgggcttctgcacatctcgtctcattgcttctggtggtggtggtggtggtggtg 780
Db 721 tgcttcaactcgggcttctgcacatctcgtctcattgcttctggtggtggtggtggtggtggtg 780
QY 781 catctggtcgaagaatgcacagatgctgctgacgggtaccacaaagcagtggtcagtgca 840
Db 781 catctggtcgaagaatgcacagatgctgctgacgggtaccacaaagcagtggtcagtgca 840
QY 841 atcataattggtcttcgacggccttctccatacaaatcgagtcctctataggaattaca 900
Db 841 atcataattggtcttcgacggccttctccatacaaatcgagtcctctataggaattaca 900
QY 901 gatctgactctatcttaagtctcgccggctatattccattcgcagacaatgacgcagtc 960
Db 901 gatctgactctatcttaagtctcgccggctatattccattcgcagacaatgacgcagtc 960
QY 961 ctctggtcgtcagtttctgaacggtccctctcatgtggcggtatcgatggcctctctc 1020
Db 961 ctctggtcgtcagtttctgaacggtccctctcatgtggcggtatcgatggcctctctc 1020
QY 1021 gctctcaacgctgtacaagagactgctgtctcactcaactggagcttgcgcgggacaat 1080
Db 1021 gctctcaacgctgtacaagagactgctgtctcactcaactggagcttgcgcgggacaat 1080
QY 1081 gggctggtatttctcactcatctcgaacgcatctcaacccgctggcaagttctctgttgg 1140
Db 1081 gggctggtatttctcactcatctcgaacgcatctcaacccgctggcaagttctctgttgg 1140
QY 1141 tctctattcggacgttgggggaattctggcacaatcggtggtatattctctaggtctcagc 1200
Db 1141 tctctattcggacgttgggggaattctggcacaatcggtggtatattctctaggtctcagc 1200
QY 1201 acagcttccaatgcttctgcaatctcgcctgtgtactccagcaactctctctctgactc 1260
Db 1201 acagcttccaatgcttctgcaatctcgcctgtgtactccagcaactctctctctgactc 1260
QY 1261 ccaatcgccctcactctctacaaaagcgagatccaaaagtcttctccgagcaactcgtgct 1320
Db 1261 ccaatcgccctcactctctacaaaagcgagatccaaaagtcttctccgagcaactcgtgct 1320
QY 1321 ttgtgtacacggcggtggaatcggttctctggtcaatgtgctagcgggtggttctcaagtc 1380
Db 1321 ttgtgtacacggcggtggaatcggttctctggtcaatgtgctagcgggtggttctcaagtc 1380
QY 1381 gtacacactggttcttctcagctcccaactgacgtgctcagcgccgctcaccacatgaa 1440
Db 1381 gtacacactggttcttctcagctcccaactgacgtgctcagcgccgctcaccacatgaa 1440

Qy 1501 gccaggagacattatccagggagccacttgagcttgacagcgggtcgctcgagcagaa 1560
|||||
Db 1501 gccaggagacattatccagggagccacttgagcttgacagcgggtcgctcgagcagaa 1560

Qy 1561 ttccaagttgggccaatga 1578
|||||

Db 1561 ttccaagttgggccaatga 1578

RESULT 3

US-09-677-682A-7
: Sequence 7, Application US/09677682A
: GENERAL INFORMATION:
: APPLICANT: Duwick, Jon
: APPLICANT: Maddox, Joyce
: APPLICANT: Gilliam, Jacob
: APPLICANT: Folkerts, Otto
: APPLICANT: Cresta, Oswald R.
: TITLE OF INVENTION: Compositions and Methods for Fumonisin
: FILE REFERENCE: Detoxification
: CURRENT APPLICATION NUMBER: US/09/677,682A
: CURRENT FILING DATE: 2000-10-02
: PRIOR APPLICATION NUMBER: 09/351,224
: PRIOR FILING DATE: 1999-07-12
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 7
: LENGTH: 1578
: TYPE: DNA
: ORGANISM: Exophiala spinifera
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (0)..(0)
: OTHER INFORMATION: permease, fully spliced cDNA
US-09-677-682A-7

Query Match 100.0%; Score 1578; DB 26; Length 1578;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atggactccagacaaagtgatagcgagaaagcgagaaagcgagacagcagaaagac 60
|||||
Db 1 atggactccagacaaagtgatagcgagaaagcgagaaagcgagacagcagaaagac 60
Qy 61 acagagacggcgagggcgagggcgagggcgagggcgagggcgagggcgagggcgag 120
|||||
Db 61 acagagacggcgagggcgagggcgagggcgagggcgagggcgagggcgagggcgag 120
Qy 121 caatttggcaccatcaccatcgctccttggccttggcttggatttgcacacagttggcgtgat 180
|||||
Db 121 caatttggcaccatcaccatcgctccttggccttggcttggatttgcacacagttggcgtgat 180
Qy 181 atctcagcagctccagctcgccctactagcggggggcgccctactctcttctacggc 240
|||||
Db 181 atctcagcagctccagctcgccctactagcggggggcgccctactctcttctacggc 240
Qy 241 atctcactcagctcctctgctacatcgatcgcttcttctcattgacggcagctgacccagc 300
|||||
Db 241 atctcactcagctcctctgctacatcgatcgcttcttctcattgacggcagctgacccagc 300
Qy 301 gtctaccgcagctggcggtggcacaatacatttgcgtcgatcctggcagcagcacaatac 360
|||||
Db 301 gtctaccgcagctggcggtggcacaatacatttgcgtcgatcctggcagcagcacaatac 360
Qy 361 aatcgagcatttcaatcgctcgagctcgctgctgctgcttcttctcattgagctgctgctc 420
|||||
Db 361 aatcgagcatttcaatcgctcgagctcgctgctgctgcttcttctcattgagctgctgctc 420
Qy 421 agctcagtgacatgatacctgctcacaagatcccgcgctgataccgctcctatagtcac 480
|||||
Db 421 agctcagtgacatgatacctgctcacaagatcccgcgctgataccgctcctatagtcac 480

Qy 481 acatactccagagattcgaggacatgctcttcctcatctacagaggagtcgcgctggtg 540
|||||
Db 481 acatactccagagattcgaggacatgctcttcctcatctacagaggagtcgcgctggtg 540
Qy 541 ctctgttcaactgtttgcccgtgaaagaaaccccttgggttcatgaaatcggattcggc 600
|||||
Db 541 ctctgttcaactgtttgcccgtgaaagaaaccccttgggttcatgaaatcggattcggc 600
Qy 601 ctcaagatcgctctctctgctgctcttcttctcgcattctagcgcgctcacaacccaaag 660
|||||
Db 601 ctcaagatcgctctctctgctgctcttcttctcgcattctagcgcgctcacaacccaaag 660
Qy 661 gctccaaactcaaggtatggactgcttggagcaactatactgctgctcgcagcggcgtc 720
|||||
Db 661 gctccaaactcaaggtatggactgcttggagcaactatactgctgctcgcagcggcgtc 720
Qy 721 tgccttcatcctggccttctcgacatcctcgtcttctcatttgcgttggagcgcagcaatg 780
|||||
Db 721 tgccttcatcctggccttctcgacatcctcgtcttctcatttgcgttggagcgcagcaatg 780
Qy 781 catctggctgaagaaatgcacagatgctcgtctacggtaccacaagcagtgctcagtcga 840
|||||
Db 781 catctggctgaagaaatgcacagatgctcgtcgtcacggtaccacaagcagtgctcagtcga 840
Qy 841 atcataattggcttctgacccgcttcttcataatacaaatcgcagttctgtatggaaatca 900
|||||
Db 841 atcataattggcttctgacccgcttcttcataatacaaatcgcagttctgtatggaaatca 900
Qy 901 gatctgactctattctaaagtctcccgagctatattccattcgcagacaaatgaacagctcc 960
|||||
Db 901 gatctgactctattctaaagtctcccgagctatattccattcgcagacaaatgaacagctcc 960
Qy 961 ctctcgctcagtttttgcacggctcctctcctcctcctcctcctcctcctcctcctcctc 1020
|||||
Db 961 ctctcgctcagtttttgcacggctcctcctcctcctcctcctcctcctcctcctcctcctc 1020
Qy 1021 gcctcacaagcgtgtacaaagagactcgctctcgtactcaactcagctgttgcgcgggacaat 1080
|||||
Db 1021 gcctcacaagcgtgtacaaagagactcgctctcgtactcaactcagctgttgcgcgggacaat 1080
Qy 1081 ggcctggtatttttccactcatctcgaacgcttccctcctcctcctcctcctcctcctcctc 1140
|||||
Db 1081 ggcctggtatttttccactcatctcgaacgcttccctcctcctcctcctcctcctcctcctc 1140
Qy 1141 tctctattcgcagctcggggaattcttgccacatcgcgatgatatattcttaggtctcagc 1200
|||||
Db 1141 tctctattcgcagctcggggaattcttgccacatcgcgatgatatattcttaggtctcagc 1200
Qy 1201 acagcttcaatgctccttgcctcctcctcctcctcctcctcctcctcctcctcctcctc 1260
|||||
Db 1201 acagcttcaatgctccttgcctcctcctcctcctcctcctcctcctcctcctcctcctc 1260
Qy 1261 ccaatcgccctactcctctacaaagcagatccaaagtcttgcgcagacactcagctgct 1320
|||||
Db 1261 ccaatcgccctactcctctacaaagcagatccaaagtcttgcgcagacactcagctgct 1320
Qy 1321 ttgtgttaccgctgggaatcggttcttggctcctcctcctcctcctcctcctcctcctcctc 1380
|||||
Db 1321 ttgtgttaccgctgggaatcggttcttggctcctcctcctcctcctcctcctcctcctcctc 1380
Qy 1381 gtacacacactgttttctcagcttcccaactgacccgtgctcagcggcgctcacaacatgaat 1440
|||||
Db 1381 gtacacacactgttttctcagcttcccaactgacccgtgctcagcggcgctcacaacatgaat 1440
Qy 1441 tacacaagtcgattatagcgttgcacttgccttctgctgcttgcctgagctggctgctgctc 1500
|||||
Db 1441 tacacaagtcgattatagcgttgcacttgccttctgctgcttgcctgagctggctgctgctc 1500
Qy 1501 gccaggagacattatcagggagcccaacttggagcttgcagggagcgggtcgctcgagcagaa 1560
|||||
Db 1501 gccaggagacattatcagggagcccaacttggagcttgcagggagcgggtcgctcgagcagaa 1560

QY 1561 ttccaagtgggccaatga 1578
Db 1561 ttccaagtgggccaatga 1578

RESULT 4
US-09-677-682B-7
; Sequence 7, Application US/09677682B
; GENERAL INFORMATION:
; APPLICANT: Duwick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 35718/204101
; CURRENT APPLICATION NUMBER: US/09/677,682B
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1578
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: permealase, fully spliced cDNA
US-09-677-682B-7

Query Match 100.0%; Score 1578; DB 26; Length 1578;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atggactcagacacaaatgatacgcgagaaagcgggacagcagacaaagaaac 60
Db 1 atggactcagacacaaatgatacgcgagaaagcgggacagcagacaaagaaac 60
QY 61 acagagacggcggcggcaggcggcggcggcggcggcggcggcggcggcggcggc 120
Db 61 acagagacggcggcggcaggcggcggcggcggcggcggcggcggcggcggcggc 120
QY 121 caatttggcaccatcaccatcgttcttcttcttcttcttcttcttcttcttctt 180
Db 121 caatttggcaccatcaccatcgttcttcttcttcttcttcttcttcttcttctt 180
QY 181 atctcaggcagctctccagctcgcctactagcgggggggggggggggggggggg 240
Db 181 atctcaggcagctctccagctcgcctactagcgggggggggggggggggggggg 240
QY 241 atctcaggcagctctcgtctacatcgtcgtcttcttcttcttcttcttcttctt 300
Db 241 atctcaggcagctctcgtctacatcgtcgtcttcttcttcttcttcttcttctt 300
QY 301 gtcacacagctcggcggcggcggcggcggcggcggcggcggcggcggcggcggc 360
Db 301 gtcacacagctcggcggcggcggcggcggcggcggcggcggcggcggcggcggc 360
QY 361 aatcaggagcatttatacagctgctgggactcgtgctgctgctgctgctgctgctg 420
Db 361 aatcaggagcatttatacagctgctgggactcgtgctgctgctgctgctgctgctg 420
QY 421 agctcaggagcatttatacagctgctgggactcgtgctgctgctgctgctgctgctg 480
Db 421 agctcaggagcatttatacagctgctgggactcgtgctgctgctgctgctgctgctg 480
QY 481 acatactccagagcttggcagctgcttcttcttcttcttcttcttcttcttcttct 540
Db 481 acatactccagagcttggcagctgcttcttcttcttcttcttcttcttcttcttct 540

QY 541 ctcttgttcaacttcttgcctgaaagaaaccccttgggttcaatgaaatcggaattcggc 600
Db 541 ctcttgttcaacttcttgcctgaaagaaaccccttgggttcaatgaaatcggaattcggc 600
QY 601 ctcaagatcgctctcttcttgccttcttcttcttcttcttcttcttcttcttcttct 660
Db 601 ctcaagatcgctctcttcttgccttcttcttcttcttcttcttcttcttcttcttct 660
QY 661 gctcaaaactcagagtgatgagctcttggagcaactatactgctgctgagcggctc 720
Db 661 gctcaaaactcagagtgatgagctcttggagcaactatactgctgctgagcggctc 720
QY 721 tgcctcatctgggcttcttgcacatctgcttcttcttcttcttcttcttcttcttct 780
Db 721 tgcctcatctgggcttcttgcacatctgcttcttcttcttcttcttcttcttcttct 780
QY 781 catctggctgaagaaatgcacagatgctgctgctgacggtaacccaaagcagtggtcagtgca 840
Db 781 catctggctgaagaaatgcacagatgctgctgctgacggtaacccaaagcagtggtcagtgca 840
QY 841 atcataattggcttcttgcacccgttcttcttcttcttcttcttcttcttcttctt 900
Db 841 atcataattggcttcttgcacccgttcttcttcttcttcttcttcttcttcttctt 900
QY 901 gatctgactctatcttaagtcttccgctgcttcttcttcttcttcttcttcttctt 960
Db 901 gatctgactctatcttaagtcttccgctgcttcttcttcttcttcttcttcttctt 960
QY 961 ctctgctgctcagtttgcacagctctcttcttcttcttcttcttcttcttcttctt 1020
Db 961 ctctgctgctcagtttgcacagctctcttcttcttcttcttcttcttcttcttctt 1020
QY 1021 gcctcaacgctgtacaagagactgctctgactcactcactcactcactcactcact 1080
Db 1021 gcctcaacgctgtacaagagactgctctgactcactcactcactcactcactcact 1080
QY 1081 gggctggtatttccactctatctgaacgcttcttcttcttcttcttcttcttctt 1140
Db 1081 gggctggtatttccactctatctgaacgcttcttcttcttcttcttcttcttctt 1140
QY 1141 tctctattcggcagcttgggaaattcttggcacaatgagatgatatattctaggtctagc 1200
Db 1141 tctctattcggcagcttgggaaattcttggcacaatgagatgatatattctaggtctagc 1200
QY 1201 acagcttccaatgcttggtaattcgcgcttcttcttcttcttcttcttcttcttct 1260
Db 1201 acagcttccaatgcttggtaattcgcgcttcttcttcttcttcttcttcttcttct 1260
QY 1261 ccaatcgccctactctcttaccacaaagcagagatccaaagtcttcttccgagcactcgtgct 1320
Db 1261 ccaatcgccctactctcttaccacaaagcagagatccaaagtcttcttccgagcactcgtgct 1320
QY 1321 ttctgttaccgcttggaaatcgggttcttcttcttcttcttcttcttcttcttctt 1380
Db 1321 ttctgttaccgcttggaaatcgggttcttcttcttcttcttcttcttcttcttctt 1380
QY 1381 gtcacacagctgttttctcagcttccactgacgctgcttcttcttcttcttcttctt 1440
Db 1381 gtcacacagctgttttctcagcttccactgacgctgcttcttcttcttcttcttctt 1440
QY 1441 tacacaagtcgattatagcgttgcacttcttcttcttcttcttcttcttcttcttct 1500
Db 1441 tacacaagtcgattatagcgttgcacttcttcttcttcttcttcttcttcttcttct 1500
QY 1501 gccaggagcattatcagggaacccacttggagcttggagcttggagcttggagcttggagcagaa 1560
Db 1501 gccaggagcattatcagggaacccacttggagcttggagcttggagcttggagcagaa 1560
QY 1561 ttccaagtgggccaatga 1578
Db 1561 ttccaagtgggccaatga 1578

; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/351,224E
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1764
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: permealase, partially spliced cDNA
US-09-351-224E-6

Query Match 100.0%; Score 1578; DB 17; Length 1764;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	atggactccagaccagtgagatcgagcggcgagaaagcggggacagggcgagacaacgagaac	60
Db	5	atggactccagaccagtgagatcgagcggcgagaaagcggggacagggcgagacaacgagaac	64
Qy	61	acagagacggcgcgcgaggggagtggtggtccgagtcctggaacgttctctggagagaaa	120
Db	65	acagagacggcgcgcgaggggagtggtggtccgagtcctggaacgttctctggagagaaa	124
Qy	121	caattggcaccatcacctgagtcctggtggtggtggtggtggtggtggtggtggtggtggt	180
Db	125	caattggcaccatcacctgagtcctggtggtggtggtggtggtggtggtggtggtggtggt	184
Qy	181	atctcaggcagtcctcagtcgctcactagcgggggggggggggggggggggggggggggggg	240
Db	185	atctcaggcagtcctcagtcgctcactagcgggggggggggggggggggggggggggggggg	244
Qy	241	atctcagtcagtcctcagtcgctcactagcgggggggggggggggggggggggggggggggg	300
Db	245	atctcagtcagtcctcagtcgctcactagcgggggggggggggggggggggggggggggggg	304
Qy	301	gtctacccgagtcggcgagcagtcctcagtcgctcactagcgggggggggggggggggggggg	360
Db	305	gtctacccgagtcggcgagcagtcctcagtcgctcactagcgggggggggggggggggggggg	364
Qy	361	aatcgagcagtcctcagtcgctcactagcgggggggggggggggggggggggggggggggg	420
Db	365	aatcgagcagtcctcagtcgctcactagcgggggggggggggggggggggggggggggggg	424
Qy	421	agtcagtcagtcctcagtcgctcactagcgggggggggggggggggggggggggggggggg	480
Db	425	agtcagtcagtcctcagtcgctcactagcgggggggggggggggggggggggggggggggg	484
Qy	481	acatactccagtcagtcgctcactagcgggggggggggggggggggggggggggggggg	540
Db	485	acatactccagtcagtcgctcactagcgggggggggggggggggggggggggggggggg	544
Qy	541	ctctgttcaactgttggcgtgagtcgctcactagcgggggggggggggggggggggggggggg	600
Db	545	ctctgttcaactgttggcgtgagtcgctcactagcgggggggggggggggggggggggggggg	604
Qy	601	ctcagtcagtcctcagtcgctcactagcgggggggggggggggggggggggggggggggg	660
Db	605	ctcagtcagtcctcagtcgctcactagcgggggggggggggggggggggggggggggggg	664
Qy	661	gtctcagtcagtcgctcagtcgctcactagcggggggggggggggggggggggggggggggg	720
Db	665	gtctcagtcagtcgctcagtcgctcactagcggggggggggggggggggggggggggggggg	724
Qy	721	tgcttcaactgttggcgtgagtcgctcactagcgggggggggggggggggggggggggggggg	780

Db	725	tgcttcaactgttggcgtgagtcgctcactagcggggggggggggggggggggggggggggggg	784
Qy	781	catctgctgagtcgctcagtcgctcactagcggggggggggggggggggggggggggggggg	840
Db	785	catctgctgagtcgctcagtcgctcactagcggggggggggggggggggggggggggggggg	844
Qy	841	atcataatgtgcttgcacgcgcttccatatacacaatcgagtcgctcagtcgctcagtcgctc	900
Db	845	atcataatgtgcttgcacgcgcttccatatacacaatcgagtcgctcagtcgctcagtcgctc	904
Qy	901	gactcgactctattcaagtcgctcagtcgctcactagcgggggggggggggggggggggggg	960
Db	905	gactcgactctattcaagtcgctcagtcgctcactagcgggggggggggggggggggggggg	964
Qy	961	cttcggtcgctcagtcgctcagtcgctcactagtcggtggtggtggtggtggtggtggtggtg	1020
Db	965	cttcggtcgctcagtcgctcagtcgctcactagtcggtggtggtggtggtggtggtggtggtg	1024
Qy	1021	gcctcacaacgctgtacaagagactgctcagtcgctcactagtcggtggtggtggtggtggtg	1080
Db	1025	gcctcacaacgctgtacaagagactgctcagtcgctcactagtcggtggtggtggtggtggtg	1084
Qy	1081	gggtcggtgcttccactcagtcgctcagtcgctcactagtcggtggtggtggtggtggtggtg	1140
Db	1085	gggtcggtgcttccactcagtcgctcagtcgctcactagtcggtggtggtggtggtggtggtg	1144
Qy	1141	tctctatcgccgactggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg	1200
Db	1145	tctctatcgccgactggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg	1204
Qy	1201	acagcttccatgctcagtcgctcagtcgctcactagtcggtggtggtggtggtggtggtggtg	1260
Db	1205	acagcttccatgctcagtcgctcagtcgctcactagtcggtggtggtggtggtggtggtggtg	1264
Qy	1261	ccaatcgccctcagtcgctcagtcgctcactagtcggtggtggtggtggtggtggtggtggtg	1320
Db	1265	ccaatcgccctcagtcgctcagtcgctcactagtcggtggtggtggtggtggtggtggtggtg	1324
Qy	1321	ttgtgttccgctggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg	1380
Db	1325	ttgtgttccgctggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg	1384
Qy	1381	gtcaccaactgcttctcagtcgctcagtcgctcactagtcggtggtggtggtggtggtggtg	1440
Db	1385	gtcaccaactgcttctcagtcgctcagtcgctcactagtcggtggtggtggtggtggtggtg	1444
Qy	1441	tacacaagtcgctcagtcgctcagtcgctcactagtcggtggtggtggtggtggtggtggtg	1500
Db	1445	tacacaagtcgctcagtcgctcagtcgctcactagtcggtggtggtggtggtggtggtggtg	1504
Qy	1501	gccaggaagcattatcaggggacccacttggagtcggtgagtcggtggtggtggtggtggtg	1560
Db	1505	gccaggaagcattatcaggggacccacttggagtcggtgagtcggtggtggtggtggtggtg	1564
Qy	1561	tttcaagtcggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg	1620
Db	1565	tttcaagtcggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg	1624

RESULT 8
US-09-677-488A-6
; Sequence 6, Application US/09677488A
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 35718/204100
; CURRENT APPLICATION NUMBER: US/09/677,488A

Mon Apr 29 09:00:43 2002

```

: CURRENT FILING DATE: 2000-10-02
: PRIOR APPLICATION NUMBER: 09/351,224
: PRIOR FILING DATE: 1999-07-12
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 6
: LENGTH: 1764
: TYPE: DNA
: ORGANISM: Exophiala spinifera
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (0)...(0)
: OTHER INFORMATION: permease, partially s
US-09-677-488A-6

```

[illegible]

```

RESULT          9
US-09-677-682A-6
: Sequence 6, Application US/09677682A
:
: GENERAL INFORMATION:
:
: APPLICANT: Duwick, Jon
:
: APPLICANT: Maddox, Joyce
:
: APPLICANT: Gilliam, Jacob
:
: APPLICANT: Folkerts, Otto
:
: APPLICANT: Grasta, Oswald R.
:
: TITLE OF INVENTION: Compositions and Methods
:
: TITLE OF INVENTION: Detoxification
:
: FILE REFERENCE: 35718/204101
:
: CURRENT APPLICATION NUMBER: US/09/677,682A
:
: CURRENT FILING DATE: 2000-10-02
:
: PRIOR APPLICATION NUMBER: 09/351,224
:
: PRIOR FILING DATE: 1999-07-12
:
: NUMBER OF SEQ ID NOS: 11
:

```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1764
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: permease, partially spliced cDNA
US-09-677-682A-6

Query Match      100.0%; Score 1578; DB 26; Length 1764;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggactcagaccagtgatagcgaggagaaagcgaggagacaaagcagacagacagaaagac 60
Db 5 atggactcagaccagtgatagcgaggagaaagcgaggagacaaagcagacagacagaaagac 64
QY 61 acagagacggcgaggcagatggtgcggtccgagtcctcctgaacgttctctctggagaagaaa 120
Db 65 acagagacggcgaggcagatggtgcggtccgagtcctcctgaacgttctctctggagaagaaa 124
QY 121 caatttggcacatcacacatcgctgcttggccttctggtgatttgcacagttgggctggt 180
Db 125 caatttggcacatcacacatcgctgcttggccttctggtgatttgcacagttgggctggt 184
QY 181 atctcagcagctccagctgcgctactagcggaggagcgccgctcaetctctctttacggc 240
Db 185 atctcagcagctccagctgcgctactagcggaggagcgccgctcaetctctctttacggc 244
QY 241 atcctaaacagatctcgtctacatctgcacatcgatcgcttctctacatagcgaactgacacgc 300
Db 245 atcctaaacagatctcgtctacatctgcacatcgatcgcttctctacatagcgaactgacacgc 304
QY 301 gtctaccgagctgcggtggccaatatcatcttggcgatcctgtgaccccaaaatcaatc 360
Db 305 gtctaccgagctgcggtggccaatatcatcttggcgatcctgtgaccccaaaatcaatc 364
QY 361 aatcgagacatttcatcgttgcgactcgctgctgttcttcttcatgagtcgctatcgga 420
Db 365 aatcgagacatttcatcgttgcgactcgctgctgttcttcttcatgagtcgctatcgga 424
QY 421 agctcagtgacatgacatctcgtctcaacagatccgcgctgacatagcgcctatagtcac 480
Db 425 agctcagtgacatgacatctcgtctcaacagatccgcgctgacatagcgcctatagtcac 484
QY 481 acatactccaggttctggtgacatgcttctctcatctacagagggagtcgctggtggtg 540
Db 485 acatactccaggttctggtgacatgcttctctcatctacagagggagtcgctggtggtg 544
QY 541 ctcttggtcaactgtttgctcctgaaagaaaccccttgggttcattgaaatcggtattcggc 600
Db 545 ctcttggtcaactgtttgctcctgaaagaaaccccttgggttcattgaaatcggtattcggc 604
QY 601 ctcaagcagctcctcttctgctgactccttcttctcgcctatagcgcgttccaaaccccaag 660
Db 605 ctcaagcagctcctcttctgctgactccttcttctcgcctatagcgcgttccaaaccccaag 664
QY 661 gctccaaactcagagatgagctgttgagcaactactgctggtgagcagggctc 720
Db 665 gctccaaactcagagatgagctgttgagcaactactgctggtgagcagggctc 724
QY 721 tgcctcactcctggcctctcgacatcctgcttcttctggttctggtcttgagcgcgcaatg 780
Db 725 tgcctcactcctggcctctcgacatcctgcttcttctggttctggtcttgagcgcgcaatg 784
QY 781 catctggctgaagaatgcacagatgctgctgacgttaccacaaagcagtggtcagtcga 840
Db 785 catctggctgaagaatgcacagatgctgctgacgttaccacaaagcagtggtcagtcga 844
QY 841 atcataatgtgctctgcgcgcttctctcatatacaaatcgagttctgtatggaattaca 900
```

```
Db 845 atcataatgtgctctgcgcgcttctctcatatacaaatcgagttctgtatggaattaca 904
QY 901 gatctcgactctattcttaagtctcgcgcgctatattccattctcgagacaatgacgcagtc 960
Db 905 gatctcgactctattcttaagtctcgcgcgctatattccattctcgagacaatgacgcagtc 964
QY 961 ctctggctcgctcagtttgcanaaggtccctctcatgtggtgggtatctggtgatggcctcttc 1020
Db 965 ctctggctcgctcagtttgcanaaggtccctctcatgtggtgggtatctggtgatggcctcttc 1024
QY 1021 gccctcaacgctgtacaagagactgctctcgaacgactcgaacgactcgaacgactcgaacgact 1080
Db 1025 gccctcaacgctgtacaagagactgctctcgaacgactcgaacgactcgaacgactcgaacgact 1084
QY 1081 ggggtggtatttccactctcgaacgactcgaacgactcgaacgactcgaacgactcgaacgact 1140
Db 1085 ggggtggtatttccactctcgaacgactcgaacgactcgaacgactcgaacgactcgaacgact 1144
QY 1141 tctctattcgcgactcgtgggaattctggcaacatgctggagatgtatatctctctggttctagc 1200
Db 1145 tctctattcgcgactcgtgggaattctggcaacatgctggagatgtatatctctctggttctagc 1204
QY 1201 acagcttctcaatgcttgggtcaattctcgcgcttgttactccagcaactctctctctgact 1260
Db 1205 acagcttctcaatgcttgggtcaattctcgcgcttgttactccagcaactctctctctgact 1264
QY 1261 ccaatcgccctactcctctacaaagcgagatccaaagttcttgcgagcaactctggtct 1320
Db 1265 ccaatcgccctactcctctacaaagcgagatccaaagttcttgcgagcaactctggtct 1324
QY 1321 ttgtgttaccgctgggaatcgggttcttggccaatgtgctagcgtgtggtctcttctacgctcc 1380
Db 1325 ttgtgttaccgctgggaatcgggttcttggccaatgtgctagcgtgtggtctcttctacgctcc 1384
QY 1381 gtccaccctgtgttttccagcttcccaactgacogtgcctacggtccgctcgaacatgaat 1440
Db 1385 gtccaccctgtgttttccagcttcccaactgacogtgcctacggtccgctcgaacatgaat 1444
QY 1441 taccaagtgagattatagcgttgcacttgccttcttgggtcttgaactgggtctggtgcat 1500
Db 1445 taccaagtgagattatagcgttgcacttgccttcttgggtcttgaactgggtctggtgcat 1504
QY 1501 gccagggaagcattatcagggaccccaacttggagacttgacggagcgggtctgctggagcagaa 1560
Db 1505 gccagggaagcattatcagggaccccaacttggagacttgacggagcgggtctgctggagcagaa 1564
QY 1561 ttccaagttggggccatga 1578
Db 1565 ttccaagttggggccatga 1582

RESULT 10
US-09-677-682B-6
; Sequence 6, Application US/09677682B
; GENERAL INFORMATION:
; APPLICANT: DuVick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Composition and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/204101
; CURRENT APPLICATION NUMBER: US/09/677,682B
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1764
; TYPE: DNA
```

```

; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: permealase, partially spliced cDNA
US-09-677-682B-6

```

Query Match		100.0%	Score 1578	DB 26	Length 1764
Best Local Similarity		100.0%	Pred. No. 0		
Matches 1578	Conservative 0	Mismatches 0	Indels 0	Gaps 0	
QY	1	atggactcagacaaatggatagcgcgagaaagcggagacagcagacaaagaagac	60		
DB	5	atggactcagacaaatggatagcgcgagaaagcggagacagcagacaaagaagac	64		
QY	61	acagagacggcgggcgaggtggcgctcgagtcctcgaacgttctctcggagagaa	120		
DB	65	acagagacggcgggcgaggtggcgctcgagtcctcgaacgttctctcggagagaa	124		
QY	121	caatttggcacatcacatcgctccttggcctttgtattgacacagttaggctggt	180		
DB	125	caatttggcacatcacatcgctccttggcctttgtattgacacagttaggctggt	184		
QY	181	atctcaggcagttctccagtcgcgcctactagcggggggggcccgtaactctcttcaagg	240		
DB	185	atctcaggcagttctccagtcgcgcctactagcggggggggcccgtaactctcttcaagg	244		
QY	241	atctcaatcagtaactctcgtctacatctgcatcgcttctctatcagcgaactgacacg	300		
DB	245	atctcaatcagtaactctcgtctacatctgcatcgcttctctatcagcgaactgacacg	304		
QY	301	gtctaccgagctgcgggtggccaatatcatctttggcgcgatactcctggcaccaaaatcaatc	360		
DB	305	gtctaccgagctgcgggtggccaatatcatctttggcgcgatactcctggcaccaaaatcaatc	364		
QY	361	aatcgagagcaattctaatcagctgtgcgagctcgtctcgttcttcaaggatcgatatcgga	420		
DB	365	aatcgagagcaattctaatcagctgtgcgagctcgtctcgttcttcaaggatcgatatcgga	424		
QY	421	agctcagtgacaatgatactctgtctcaacagatacccggcgtgatagcgcgcatactagtcac	480		
DB	425	agctcagtgacaatgatactctgtctcaacagatacccggcgtgatagcgcgcatactagtcac	484		
QY	481	acatactccaggaattcgtagcatgtctctctcatctacgagggagtcgcgcgttggtgtg	540		
DB	485	acatactccaggaattcgtagcatgtctctctcatctacgagggagtcgcgcgttggtgtg	544		
QY	541	ctcttgctcaactgtttgcctcgaaagaaaccccttggttctatgaaatcggatctcggc	600		
DB	545	ctcttgctcaactgtttgcctcgaaagaaaccccttggttctatgaaatcggatctcggc	604		
QY	601	ctcagagatgcctctcttgtagatactcttatagcgcattctagcggcggtccaaaccccaag	660		
DB	605	ctcagagatgcctctcttgtagatactcttatagcgcattctagcggcggtccaaaccccaag	664		
QY	661	gtctcaaaactcacaggtatggactgcttgagacaatactatggctggtgcgcgcgcgc	720		
DB	665	gtctcaaaactcacaggtatggactgcttgagacaatactatggctggtgcgcgcgcgc	724		
QY	721	tgcttcaactcggcctttgcacatcgtgcttcaatttcattggcttgagcgcgacaatg	780		
DB	725	tgcttcaactcggcctttgcacatcgtgcttcaatttcattggcttgagcgcgacaatg	784		
QY	781	catctggctggaagaatcacagatgctcgtacaggtaccaccaagcagtggtcaatgca	840		
DB	785	catctggctggaagaatcacagatgctcgtacaggtaccaccaagcagtggtcaatgca	844		
QY	841	atcataattggcctctgcacgcgctttccatatacaaatcgagttctctgtatggaattaca	900		
DB	845	atcataattggcctctgcacgcgctttccatatacaaatcgagttctctgtatggaattaca	904		
QY	901	gatctcgactctatctcaagtctccgcgcgctaattctccatttcgagacaatagaacgattcc	960		

```

RESULT 11
US-09-882-694-6
: Sequence 6, Application US/09882694
: GENERAL INFORMATION:
: APPLICANT: Duwick, Jon
: APPLICANT: Maddox, Joyce
: APPLICANT: Gilliam, Jacob
: APPLICANT: Folkerts, Otto
: APPLICANT: Crasta, Oswald R.
: TITLE OF INVENTION: Compositions and Methods for Fumonisin
: TITLE OF INVENTION: Detoxification
: FILE REFERENCE: 35718/208255
: CURRENT APPLICATION NUMBER: US/09/882,694
: CURRENT FILING DATE: 2001-06-12
: PRIOR APPLICATION NUMBER: 09/351,224
: PRIOR FILING DATE: 1999-07-12
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 6
: LENGTH: 1764
: TYPE: DNA
: ORGANISM: Exophiala spinifera
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (0)...(0)

```

! OTHER INFORMATION: permealase, partially spliced cDNA
US-09-882-694-6

Query Match 100.0%; Score 1578; DB 33; Length 1764;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgactccagacacagatgatacggcagagaaagcggagacaaacagacaaac 60
Db 5 atgactccagacacagatgatacggcagagaaagcggagacaaacagacaaac 64

QY 61 acagagacgagcggcggcagagtgctgcgtccagatccctgaacgttctcttgagagaa 120
Db 65 acagagacgagcggcggcagagtgctgcgtccagatccctgaacgttctcttgagagaa 124

QY 121 caatttggcaccatcacatcgctgctctggccttggcttggatttgaacagttggctgggt 180
Db 125 caatttggcaccatcacatcgctgctctggccttggcttggatttgaacagttggctgggt 184

QY 181 atctcagagcagctcagctcagctcagctcagctcagctcagctcagctcagctcagctc 240
Db 185 atctcagagcagctcagctcagctcagctcagctcagctcagctcagctcagctcagctc 244

QY 241 atctcagagcagctcagctcagctcagctcagctcagctcagctcagctcagctcagctc 300
Db 245 atctcagagcagctcagctcagctcagctcagctcagctcagctcagctcagctcagctc 304

QY 301 gctcaccgagctcggcggcgaataatcttctgctcagctcagctcagctcagctcagctc 360
Db 305 gctcaccgagctcggcggcgaataatcttctgctcagctcagctcagctcagctcagctc 364

QY 361 aatcggagcagctcagctcagctcagctcagctcagctcagctcagctcagctcagctc 420
Db 365 aatcggagcagctcagctcagctcagctcagctcagctcagctcagctcagctcagctc 424

QY 421 agctcagtgacacatgatacctgctcagacagatcccgcgctgatacgcgcctatagtcac 480
Db 425 agctcagtgacacatgatacctgctcagacagatcccgcgctgatacgcgcctatagtcac 484

QY 481 acatactccagagatctgagcagtgcttctcctcagctcagctcagctcagctcagctc 540
Db 485 acatactccagagatctgagcagtgcttctcctcagctcagctcagctcagctcagctc 544

QY 541 ctctgtcagcagcttctcctcgaagaaaccccttggtctcagctcagctcagctcagctc 600
Db 545 ctctgtcagcagcttctcctcgaagaaaccccttggtctcagctcagctcagctcagctc 604

QY 601 ctacagatcgtctctctgagctcctcttctatcgcgaatctagcgcggtcccaaccccaag 660
Db 605 ctacagatcgtctctctgagctcctcttctatcgcgaatctagcgcggtcccaaccccaag 664

QY 661 gctccaaactcacaggtatgagctgcttgagcaactatctgctggtcgcagcgcgcgc 720
Db 665 gctccaaactcacaggtatgagctgcttgagcaactatctgctggtcgcagcgcgcgc 724

QY 721 tgcctcagctcagctcagctcagctcagctcagctcagctcagctcagctcagctcagctc 780
Db 725 tgcctcagctcagctcagctcagctcagctcagctcagctcagctcagctcagctcagctc 784

QY 781 catctgctgaagaaatgacagatgctcgtcagctcagctcagctcagctcagctcagctc 840
Db 785 catctgctgaagaaatgacagatgctcgtcagctcagctcagctcagctcagctcagctc 844

QY 841 atcataattgctctcgcacgccttccatatacaatcgcagctcagctcagctcagctcagctc 900
Db 845 atcataattgctctcgcacgccttccatatacaatcgcagctcagctcagctcagctcagctc 904

QY 901 gatctcagctcagctcagctcagctcagctcagctcagctcagctcagctcagctcagctc 960
Db 905 gatctcagctcagctcagctcagctcagctcagctcagctcagctcagctcagctcagctc 964

QY 961 ctctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1020

Db 965 ctctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1024

QY 1021 gccctcaacgctgtacaagagactgctcgtcagctcacctcagcttgcggagcaaat 1080

Db 1025 gccctcaacgctgtacaagagactgctcgtcagctcacctcagcttgcggagcaaat 1084

QY 1081 gggctcggctatttccactcactcagacgacatcccgctcggcagcttccgtttgg 1140

Db 1085 gggctcggctatttccactcactcagacgacatcccgctcggcagcttccgtttgg 1144

QY 1141 tctctattcggagacactggggaattctcggcacaatgggagatgatatattttaggttctagc 1200

Db 1145 tctctattcggagacactggggaattctcggcacaatgggagatgatatattttaggttctagc 1204

QY 1201 acagctttcaatgcttggcttggcaattcccgcttggctcagcagcaactcctctcctgact 1260

Db 1205 acagctttcaatgcttggcttggcaattcccgcttggctcagcagcaactcctctcctgact 1264

QY 1261 ccaatcgccctcactcctctaccaaaagcagatcccaagtctcttgcgagcactcgtgct 1320

Db 1265 ccaatcgccctcactcctctaccaaaagcagatcccaagtctcttgcgagcactcgtgct 1324

QY 1321 ttgtgttaccgctggaatcggttctctggtcaatgtcgtcagcggcgttctcagctcc 1380

Db 1325 ttgtgttaccgctggaatcggttctctggtcaatgtcgtcagcggcgttctcagctcc 1384

QY 1381 gtcaccactgtgttttccactcctccactgacccgtgctcagcgcgcgtcaccactgact 1440

Db 1385 gtcaccactgtgttttccactcctccactgacccgtgctcagcgcgcgtcaccactgact 1444

QY 1441 tacacaagtgcgactatagcgttgcacttgccttcttgcgttgcagctcggctcgtgcat 1500

Db 1445 tacacaagtgcgactatagcgttgcacttgccttcttgcgttgcagctcggctcgtgcat 1504

QY 1501 gccagggaagcattatcaggacccacttggacttgcagcggcggcgtcgcggagcagaa 1560

Db 1505 gccagggaagcattatcaggacccacttggacttgcagcggcggcgtcgcggagcagaa 1564

QY 1561 ttccaagtgtggccatga 1578

Db 1565 ttccaagtgtggccatga 1582

RESULT 12
US-09-882-694A-6
; Sequence 6, Application US/09882694A
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/882,694A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1764
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)-(0)
; OTHER INFORMATION: permealase, partially spliced cDNA
US-09-882-694A-6

Db 301 gtctaccgactgocggtggccaatatcatatttgcgtcgatcctggcaccacaaatcaatc 360
Qy 361 aatcggagcatttcaatcgtgtcggaactcgtgtgtgtttcattggtcgtctatcgga 420
Db 361 aatcggagcatttcaatcgtgtcggaactcgtgtgtgtttcattggtcgtctatcgga 420
Qy 421 agctcagtgacatgatactctcacaagatcccgccgctgataccgcttatagtcac 480
Db 421 agctcagtgacatgatactctcacaagatcccgccgctgataccgcttatagtcac 480
Qy 481 acatactccaggattcgtggaatgtcttctcctcatcagaagagtcgctgtgtgtg 540
Db 481 acatactccaggattcgtggaatgtcttctcctcatcagaagagtcgctgtgtgtg 540
Qy 541 ctcttgctcaactgtttgcctgaaagaaaccccttgggttcatgaaatcgattcggc 600
Db 541 ctcttgctcaactgtttgcctgaaagaaaccccttgggttcatgaaatcgattcggc 600
Qy 601 ctccagatcgtctctcgtgatactcttcttccatctagcgcgttccaaacccaag 660
Db 601 ctccagatcgtctctcgtgatactcttcttccatctagcgcgttccaaacccaag 660
Qy 661 gctccaaactcacaggatgactgcttgagcaactatactgctggtccgacggtc 720
Db 661 gctccaaactcacaggatgactgcttgagcaactatactgctggtccgacggtc 720
Qy 721 tgcctcatcctggccttgacatcctgcttctcattgcttgcgttgacacgacatg 780
Db 721 tgcctcatcctggccttgacatcctgcttctcattgcttgcgttgacacgacatg 780
Qy 781 catctggctgaagaatgcacagatgctcgtctgacgttaccacaaagcagtgatgca 840
Db 781 catctggctgaagaatgcacagatgctcgtctgacgttaccacaaagcagtgatgca 840
Qy 841 atcataatggctctgcacgccttccatatacaatgcagcttctgtatgaaattaca 900
Db 841 atcataatggctctgcacgccttccatatacaatgcagcttctgtatgaaattaca 900
Qy 901 gatctcgactatttcaagtccgcggtatattccattccattccagacaaatgcagctc 960
Db 901 gatctcgactatttcaagtccgcggtatattccattccattccagacaaatgcagctc 960
Qy 961 ctctcgtcgtcagtttgcacagctctcctcctcgttgcgtggtatgctgctcttctc 1020
Db 961 gctcgaattcggcttttgcacagctctcctcctcgttgcgtggtatgctgctcttctc 1020
Qy 1021 gccctcaacgctgtacagagactcgtctcgtactcacctggagcttcccggaacaa 1080
Db 1021 gccctcaacgctgtacagagactcgtctcgtactcacctggagcttcccggaacaa 1080
Qy 1081 gggctgtgtatttccactatctcgaacgcatctccatcccgctggcgaagtctcgtgtg 1140
Db 1081 gggctgtgtatttccactatctcgaacgcatctccatcccgctggcgaagtctcgtgtg 1140
Qy 1141 tctctatcgcgacgtgggaattctcgcacatcggaatgataatttctaggtctagc 1200
Db 1141 tctctatcgcgacgtgggaattctcgcacatcggaatgataatttctaggtctagc 1200
Qy 1201 acaggtttcaatgctgtgtaattccgcggttgactccagaaatctccttctgac 1260
Db 1201 acaggtttcaatgctgtgtaattccgcggttgactccagaaatctccttctgac 1260
Qy 1261 ccaatcgcctactcctctacacaaagcgagatccaaagtcttctgcgagacactggt 1320
Db 1261 ccaatcgcctactcctctacacaaagcgagatccaaagtcttctgcgagacactggt 1320
Qy 1321 ttgtgttaccgctggaatcgtgttctgttcaatgtctagcgtgtcttctcagctcc 1380
Db 1321 ttgtgttaccgctggaatcgtgttctgttcaatgtctagcgtgtcttctcagctcc 1380
Qy 1381 gtcaacactgtgttttcaagtctccactgacctgacctgacctgacctgacctgacct 1440
Db 1381 gtcaacactgtgttttcaagtctccactgacctgacctgacctgacctgacctgacct 1440

Qy 1441 tacacaagtcgattatagcgttgactgtctcttgcgttgcgttgcgttgcgttgcgt 1500
Db 1441 tacacaagtcgattatagcgttgactgtctcttgcgttgcgttgcgttgcgttgcgt 1500
Qy 1501 gccagaagcattatcagagaccacacttgagcttgagcttgagcgggtcgctcgagagaa 1560
Db 1501 gccagaagcattatcagagaccacacttgagcttgagcttgagcgggtcgctcgagagaa 1560
Qy 1561 ttccaagttgggcatga 1578
Db 1561 ttccaagttgggcatga 1578

RESULT 15
US-09-677-488-7
: Sequence 7, Application US/09677488
: GENERAL INFORMATION:
: APPLICANT: DUVICK, Jon
: TITLE OF INVENTION: Compositions and Methods for Fummonisin Detoxification
: FILE REFERENCE: 35718/204100 (5718-111A)
: CURRENT APPLICATION NUMBER: US/09/677,488
: PRIOR FILING DATE: 2000-10-02
: PRIOR APPLICATION NUMBER: 09/351,224
: PRIOR FILING DATE: 1999-07-12
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO: 7
: LENGTH: 1578
: TYPE: DNA
: ORGANISM: Exophiala spinifera
: FEATURE:
: OTHER INFORMATION: permease, fully spliced cDNA
US-09-677-488-7

Query Match 98.6%; Score 1555.6; DB 26; Length 1578;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1564; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 1 atggactcagaccacaaagtgtacagcagaaagcgggacagcagacacgaagaac 60
Db 1 atggactcagaccacaaagtgtacagcagaaagcgggacagcagacacgaagaac 60
Qy 61 acagagacggcggcaggtgtgctgcagctccctgaacgttccctctgagagaaa 120
Db 61 acagagacggcggcaggtgtgctgcagctccctgaacgttccctctgagagaaa 120
Qy 121 caatttggcaccatacaccatcgtctccttggccttgcgttgcgttgcgttgcgttgcgt 180
Db 121 caatttggcaccatacaccatcgtctccttggccttgcgttgcgttgcgttgcgttgcgt 180
Qy 181 atctcaggcagctccagctcgcctactagcggggggggcgcctcactctcttaccgc 240
Db 181 atctcaggcagctccagctcgcctactagcggggggggcgcctcactctcttaccgc 240
Qy 241 atctcaggcagctcgcctactagcggggggggcgcctcactctcttaccgc 300
Db 241 atctcaggcagctcgcctactagcggggggggcgcctcactctcttaccgc 300
Qy 301 gtctaccgactcgcgttgcccaatatacttgcgtcgatcctggaccacaaatcaatc 360
Db 301 gtctaccgactcgcgttgcccaatatacttgcgtcgatcctggaccacaaatcaatc 360
Qy 361 aatcggagcatttcaacgtgtgcgacactgctgttgcgttgcgttgcgttgcgttgcgt 420
Db 361 aatcggagcatttcaacgtgtgcgacactgctgttgcgttgcgttgcgttgcgttgcgt 420
Qy 421 agctcagtgacatgatactctcacaagatcccgccgctgataccgcttatagtcac 480
Db 421 agctcagtgacatgatactctcacaagatcccgccgctgataccgcttatagtcac 480
Qy 481 acatactccaggattcgtggaatgtcttctcctcatcagaagagtcgctgtgtgtg 540

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 01:35:42 ; Search time 150.7 Seconds
(without alignments)
2572.062 Million cell updates/sec

Title: US-09-882-694A-7

Perfect score: 1578

Sequence: 1 atgagctccagaccagtggtg.....aatttcaagtgggccatga 1578

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58	3.7	7218	1	US-08-232-463-14
2	35.8	2.3	289	4	US-09-007-005-17
3	35.8	2.3	289	4	US-09-244-796-17
4	34	2.2	1473	2	US-08-541-033A-25
5	34	2.2	1473	2	US-08-828-451-25
6	34	2.2	1506	2	US-08-541-033A-23
7	34	2.2	1506	2	US-08-828-451-23
8	34	2.2	1969	2	US-08-541-033A-7
9	34	2.2	1969	2	US-08-828-451-7
10	34	2.2	2096	2	US-08-541-033A-19
11	34	2.2	2096	2	US-08-828-451-19
12	34	2.2	2099	2	US-08-541-033A-3
13	34	2.2	2099	2	US-08-828-451-3
14	34	2.2	2137	2	US-08-541-033A-18
15	34	2.2	2137	2	US-08-828-451-18
16	34	2.2	2140	2	US-08-541-033A-1
17	34	2.2	2140	2	US-08-828-451-1
18	32.8	2.1	642	4	US-09-199-637A-26
19	32.8	2.1	957	4	US-09-199-637A-22
20	32.8	2.1	1407	4	US-09-199-637A-28
21	32.8	2.1	3141	4	US-09-199-637A-12
22	32.8	2.1	42235	4	US-09-199-637A-1
23	32.8	2.1	686	4	US-09-328-111-364
24	32.2	2.0	1249	3	US-09-113-536-1
25	32.2	2.0	1249	5	PCT-US95-05785-1
26	31.6	2.0	2102	1	US-08-784-651-5
27	31.2	2.0	44377	2	US-08-804-227C-7

c	28	31.2	2.0	44377	2	US-08-804-198-1	Sequence 1, Appl
c	29	31	2.0	192	1	US-08-030-731A-45	Sequence 45, Appl
c	30	31	2.0	192	1	US-08-030-731A-46	Sequence 46, Appl
c	31	30.4	1.9	1279	1	US-08-146-010A-4	Sequence 4, Appl
c	32	30.4	1.9	1279	1	US-08-674-188-9	Sequence 9, Appl
c	33	30.4	1.9	12047	2	US-09-022-461-1	Sequence 1, Appl
c	34	30	1.9	702	4	US-08-843-417-3	Sequence 3, Appl
c	35	30	1.9	947	1	US-08-474-177-36	Sequence 36, Appl
c	36	30	1.9	947	1	US-08-487-033-36	Sequence 36, Appl
c	37	30	1.9	947	1	US-08-480-810-36	Sequence 36, Appl
c	38	30	1.9	947	1	US-08-508-735-36	Sequence 36, Appl
c	39	30	1.9	947	2	US-08-848-251-36	Sequence 36, Appl
c	40	30	1.9	947	2	US-08-486-047-36	Sequence 36, Appl
c	41	30	1.9	947	3	US-09-120-130-36	Sequence 36, Appl
c	42	30	1.9	947	3	US-09-115-252-36	Sequence 36, Appl
c	43	30	1.9	947	3	US-08-986-515-36	Sequence 36, Appl
c	44	30	1.9	947	3	US-09-120-128-36	Sequence 36, Appl
c	45	30	1.9	947	4	US-09-120-129-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14

```
Query Match          3.7%; Score 58; DB 1; Length 7218;
Best Local Similarity 3.4%; Pred. No. 2.6e-09;
Matches 13; Conservative 224; Mismatches 149; Indels 0; Gaps 0;

QY 900 agatctcgactctattctaaagttccgcggctatatctccatctcgagacaatgacgcagtc 959
   || | ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1057 AGCTTGGATYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 1116

QY 960 ccttcggtcgctcagtttgcaacggtcctctcatgtggtggtatcggtatcggtcctctt 1019
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1117 YVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 1176

QY 1020 cgcctcaacgctgtacaagagactcgtctcgactcacctcagctgagctttgcccgggacaa 1079
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1177 YVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 1236

QY 1080 tgggctgggtattttccacatctcgtgaacgcatctcccgctggaagttcctgtttg 1139
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1237 YVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 1296

QY 1140 gtctctattcgacgtgggaattctgcccacatgcggtatgtatttctaggttctag 1199
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1297 YVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 1356

QY 1200 cacagtttcaatgccttggtcaattccgctgtgtactccagcaactctcctctgat 1259
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1357 YVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 1416

QY 1260 cccaatcgccctactctctctacaaa 1285
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1417 YVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 1442

RESULT 2
US-09-007-005-17/c
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szoostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; EARLIER FILING DATE: 1998-01-14
; EARLIER FILING DATE: 1997-01-27
; EARLIER FILING DATE: 1997-01-27
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17
```

```
Query Match          2.3%; Score 35.8; DB 4; Length 289;
Best Local Similarity 5.3%; Pred. No. 0.017;
Matches 10; Conservative 83; Mismatches 94; Indels 0; Gaps 0;

QY 146 ccttgccctttgtgattgcaacagttgggtggtatctccagcagctccagctcgccc 205
   : | : | : | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 231 YCAVGYCVTYGVSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYN 172
```

4 - ' 4

```
QY 206 tactagcggggggggcccgctcaactctctcttaccgcatcctcaatcagctactctcgtctaca 265
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 171 YSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYN 112

QY 266 tctgcatcgcttctcattagccgaactgaccagcagctaccgccgactgcccgtggcccaat 325
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 111 YSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYN 52

QY 326 atcattt 332
   | : : | :
Db 51 YTYGYTY 45

RESULT 3
US-09-244-796-17/c
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szoostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; EARLIER FILING DATE: 1999-02-05
; EARLIER FILING DATE: 1997-01-27
; EARLIER FILING DATE: 1997-01-27
; EARLIER FILING DATE: 1997-11-06
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match          2.3%; Score 35.8; DB 4; Length 289;
Best Local Similarity 5.3%; Pred. No. 0.017;
Matches 10; Conservative 83; Mismatches 94; Indels 0; Gaps 0;

QY 146 ccttgccctttgtgattgcaacagttgggtggtatctccagcagctccagctcgccc 205
   : | : | : | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 231 YCAVGYCVTYGVSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYN 172

QY 206 tactagcggggggggcccgctcaactctctcttaccgcatcctcaatcagctactcgtctaca 265
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 171 YSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYN 112

QY 266 tctgcatcgcttctcattagccgaactgaccagcagctaccgccgactgcccgtggcccaat 325
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 111 YSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYN 52

QY 326 atcattt 332
   | : : | :
Db 51 YTYGYTY 45

RESULT 4
US-08-541-033A-25/c
; Sequence 25, Application US/08541033A
; Patent No. 5679941
; GENERAL INFORMATION:
```



```
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1506 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4..1464
; US-08-541-033A-23

Query Match      2.2%; Score 34; DB 2; Length 1506;
Best Local Similarity 56.1%; Pred. No. 0.24;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Oy 1460 gcgttcacattgctcttggtcttgaactgggtcgatgccaggaagcattatcagg 1519
      ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 539 GCCTCGCTCTTGGGCTTGGGCTGGAAGTCGGAGCGCCCTTGGCGCCGCCATGGGCAGG 480
      || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Oy 1520 gacccacttgagcttgacggagcgggtcgctcgagcagaaatttcaagttgggc 1573
      || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 479 GTGTCAGGCTGTCTTGAAGATCTGCTCAAAGGCAAGCAACTTCATGATGGAC 426
      || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
US-08-541-033A-7/c
; Sequence 7, Application US/08541033A
; Patent No. 5879941
; GENERAL INFORMATION:
; APPLICANT: Schmidt, Robert R.
; APPLICANT: Miller, Philip
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
; TITLE OF INVENTION: RELATING TO THE '-' AND -SUBUNITS OF GLUTAMATE
; TITLE OF INVENTION: DEHYDROGENASES AND METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/541,033A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1969 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-541-033A-7

Query Match      2.2%; Score 34; DB 2; Length 1969;
Best Local Similarity 56.1%; Pred. No. 0.3;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Oy 1460 gcgttcacattgctcttggtcttgaactgggtcgatgccaggaagcattatcagg 1519
      ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 515 GCCTCGCTCTTGGGCTTGGGCTGGAAGTCGGAGCGCCCTTGGCGCCGCCATGGGCAGG 456
      || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```


Qy 1520 gacccacttgagcttgacggacggtcgtcggagcagaattttcaagtgggc 1573
+ + + + + + + + + + + + + + + + + + +
Db 455 GTGTCAGGTCTTCAAGATCTGCTCAAAGCAAGGAACCTTCATGATGGAC 402

```

1  RESULT 9
2  US-08-828-451-7/C
3  ; Sequence 7, Application US/08828451
4  ; Patent No. 5985634
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Schmidt, Robert R.
7  ; APPLICANT: Miller, Philip
8  ; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
9  ; TITLE OF INVENTION: RELATING TO THE - AND -SUBUNITS OF GLUTAMATE
10 ; TITLE OF INVENTION: DEHYDRATION: DEHYDROGENASES AND METHODS OF USE
11 ; NUMBER OF SEQUENCES: 26
12 ; CORRESPONDENCE ADDRESS:
13 ; ADDRESSEE: Saliwanchik & Saliwanchik
14 ; STREET: 2421 N.W. 41st Street, Suite A-1
15 ; CITY: Gainesville
16 ; STATE: Florida
17 ; COUNTRY: USA
18 ; ZIP: 32606
19 ; COMPUTER READABLE FORM: disk
20 ; MEDIUM TYPE: Floppy disk
21 ; COMPUTER: IBM PC compatible
22 ; OPERATING SYSTEM: PC-DOS/MS-DOS
23 ; SOFTWARE: Patent In Release #1.0, Version #1.25
24 ; CURRENT APPLICATION DATA:
25 ; APPLICATION NUMBER: US/08/828,451
26 ; FILING DATE:
27 ; CLASSIFICATION:
28 ; PRIOR APPLICATION DATA:
29 ; APPLICATION NUMBER: US/08/541,033
30 ; FILING DATE:
31 ; ATTORNEY/AGENT INFORMATION:
32 ; NAME: Whitlock, Ted W.
33 ; REGISTRATION NUMBER: 36,965
34 ; REFERENCE/DOCKET NUMBER: UF155
35 ; TELECOMMUNICATION INFORMATION:
36 ; TELEPHONE: (904) 375-8100
37 ; TELEFAX: (904) 372-5800
38 ; INFORMATION FOR SEQ ID NO: 7:
39 ; SEQUENCE CHARACTERISTICS:
40 ; LENGTH: 1969 base pairs
41 ; TYPE: nucleic acid
42 ; STRANDEDNESS: double
43 ; TOPOLOGY: linear
44 ; MOLECULE TYPE: CDNA
45 ; US-08-828-451-7

```

RESULT 10
US-08-541-033A-19/c
; Sequence 19, Application US/08541033A
; Patent No. 5879941
; GENERAL INFORMATION:
; APPLICANT: Schmidt, Robert R.
; APPLICANT: Miller, Philip

```

1 TITLE OF INVENTION: NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
2
3 TITLE OF INVENTION: RELATING TO THE -- AND --SUBUNITS OF GLUTAMATE
4
5 TITLE OF INVENTION: DEHYDROGENASES AND METHODS OF USE
6
7 NUMBER OF SEQUENCES: 26
8
9 CORRESPONDENCE ADDRESS:
10
11 ADDRESSEE: Saliwanchik & Saliwanchik
12
13 STREET: 2421 N.W. 41st Street, Suite A-1
14
15 CITY: Gainesville
16
17 STATE: Florida
18
19 COUNTRY: USA
20
21 ZIP: 32606
22
23 COMPUTER READABLE FORM:
24
25 MEDIUM TYPE: Floppy disk
26
27 COMPUTER: IBM PC Compatible
28
29 OPERATING SYSTEM: PC-DOS/MS-DOS
30
31 SOFTWARE: Patencin Release #1.0, Version #1.25
32
33 CURRENT APPLICATION DATA:
34
35 APPLICATION NUMBER: US/08/541.033A
36
37 FILING DATE:
38
39 CLASSIFICATION: 435
40
41 ATTORNEY/AGENT INFORMATION:
42
43 NAME: Whitlock, Ted W.
44
45 REGISTRATION NUMBER: 36,965
46
47 REFERENCE/DOCKET NUMBER: UFI15
48
49 TELECOMMUNICATION INFORMATION:
50
51 TELEPHONE: (904) 375-8100
52
53 TELEFAX: (904) 372-5800
54
55 INFORMATION FOR SEQ ID NO: 19:
56
57 SEQUENCE CHARACTERISTICS:
58
59 LENGTH: 2096 base pairs
60
61 TYPE: nucleic acid
62
63 STRANDEDNESS: double
64
65 TOPOLOGY: linear
66
67 MOLECULE TYPE: CDNA
68
69 US-08-541-033A-19
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
96
```

```

RESULT 11
US-08-828-451-19/c
; Sequence 19, Application US/08028451
; Patent No. 5985634
; GENERAL INFORMATION:
; APPLICANT: Schmidt, Robert R.
; APPLICANT: Miller, Philip
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
; TITLE OF INVENTION: RELATING TO THE - AND -SUBUNITS OF GLUTAMATE
; TITLE OF INVENTION: DEHYDROGENASES AND METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/828,451
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/541,033
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Whitlock, Ted W.
;; REGISTRATION NUMBER: 36,965
;; REFERENCE/DOCKET NUMBER: UF155
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (904) 375-8100
;; TELEFAX: (904) 372-5800
;; INFORMATION FOR SEQ ID NO: 19:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2096 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: CDNA
US-08-828-451-19

Query Match 2.2%; Score 34; DB 2; Length 2096;
Best Local Similarity 56.1%; Pred. No. 0.31;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 1460 gcgttgcaacttctcttggcttgactgggtcgctgcatgccaggagcattatcagg 1519
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 640 GCGTCGCTCTTGGCCCTTGGGTCGAAGTCGGAGCGCCCTTGGCGGCCCATGGGCAGG 581
QY 1520 gaccccaactggagcttgaacgagggctgctcgagcagagaatttcagttgggc 1573
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 580 GTGGTCAGGCTGTTCTTGAGATCTGCTCAAGGCAAGGAACTTCATGATGGAC 527

RESULT 12
US-08-541-033A-3/c
; Sequence 3, Application US/08541033A
; Patent No. 5879941
; GENERAL INFORMATION:
; APPLICANT: Schmidt, Robert R.
; APPLICANT: Miller, Phillip
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
; TITLE OF INVENTION: RELATING TO THE - AND -SUBUNITS OF GLUTAMATE
; TITLE OF INVENTION: DEHYDROGENASES AND METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/541,033A
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:

;; LENGTH: 2099 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: CDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 33..1568
US-08-541-033A-3

Query Match 2.2%; Score 34; DB 2; Length 2099;
Best Local Similarity 56.1%; Pred. No. 0.31;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 1460 gcgttgcaacttctcttggcttgactgggtcgctgcatgccaggagcattatcagg 1519
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 643 GCGTCGCTCTTGGCCCTTGGGTCGAAGTCGGAGCGCCCTTGGCGGCCCATGGGCAGG 584
QY 1520 gaccccaactggagcttgaacgagggctgctcgagcagagaatttcagttgggc 1573
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 583 GTGGTCAGGCTGTTCTTGAGATCTGCTCAAGGCAAGGAACTTCATGATGGAC 530

RESULT 13
US-08-828-451-3/c
; Sequence 3, Application US/08828451
; Patent No. 5985634
; GENERAL INFORMATION:
; APPLICANT: Schmidt, Robert R.
; APPLICANT: Miller, Phillip
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
; TITLE OF INVENTION: RELATING TO THE - AND -SUBUNITS OF GLUTAMATE
; TITLE OF INVENTION: DEHYDROGENASES AND METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828,451
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/541,033
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2099 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 33..1568
US-08-828-451-3

```
Query Match          2.2%; Score 34; DB 2; Length 2099;
Best Local Similarity 56.1%; Pred. No. 0.31;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1460 qcgttgacactgctctgtggtgctgaactgggtcgatgcagcagaagcattatcagg 1519
      ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 643 GCGTCGCTCTTGCCCTTGGGTCGGAAGTCGGAGCGCCCTTGCGCGCCCATGGGCAGG 584

QY 1520 gaccccaacttgaggtgacagcagggtcgctcgagagcagaatttcaagttgggc 1573
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 583 GTGGTCAGGCTGTCTTGAAGATCTGCTCAAAGGCAAGGAACCTTCATGATGGAC 530

RESULT 14
US-08-541-033A-18/c
; Sequence 18, Application US/08541033A
; Patent No. 5879941
; GENERAL INFORMATION:
; APPLICANT: Schmidt, Robert R.
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
; TITLE OF INVENTION: RELATING TO THE - AND -SUBUNITS OF GLUTAMATE
; TITLE OF INVENTION: DEHYDROGENASES AND METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/541,033A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/541,033
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2137 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-541-033A-18

Query Match          2.2%; Score 34; DB 2; Length 2137;
Best Local Similarity 56.1%; Pred. No. 0.32;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1460 qcgttgacactgctctgtggtgctgaactgggtcgatgcagcagaagcattatcagg 1519
      ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 682 GCGTCGCTCTTGCCCTTGGGTCGGAAGTCGGAGCGCCCTTGCGCGCCCATGGGCAGG 623

QY 1520 gaccccaacttgaggtgacagcagggtcgctcgagagcagaatttcaagttgggc 1573
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 622 GTGGTCAGGCTGTCTTGAAGATCTGCTCAAAGGCAAGGAACCTTCATGATGGAC 569

RESULT 15
US-08-828-451-16/c
```

```
; Sequence 18, Application US/08828451
; Patent No. 5985634
; GENERAL INFORMATION:
; APPLICANT: Schmidt, Robert R.
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
; TITLE OF INVENTION: RELATING TO THE - AND -SUBUNITS OF GLUTAMATE
; TITLE OF INVENTION: DEHYDROGENASES AND METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828,451
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/541,033
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2137 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-828-451-16

Query Match          2.2%; Score 34; DB 2; Length 2137;
Best Local Similarity 56.1%; Pred. No. 0.32;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1460 qcgttgacactgctctgtggtgctgaactgggtcgatgcagcagaagcattatcagg 1519
      ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 682 GCGTCGCTCTTGCCCTTGGGTCGGAAGTCGGAGCGCCCTTGCGCGCCCATGGGCAGG 623

QY 1520 gaccccaacttgaggtgacagcagggtcgctcgagagcagaatttcaagttgggc 1573
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 622 GTGGTCAGGCTGTCTTGAAGATCTGCTCAAAGGCAAGGAACCTTCATGATGGAC 569

Search completed: April 27, 2002, 01:35:58
Job time: 22247 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 00:40:59 ; Search time 706.01 seconds
(without alignments)
3837.468 Million cell updates/sec

Title: US-09-882-694A-7

Perfect score: 1578

Sequence: 1 atgagtcacagaccagtggtg.....aatttcaggtggccatga 1578

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802.*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|----------------------|
| 1 | 1355.6 | 98.6 | 1764 | 22 | Exophiala spinifera |
| 2 | 46.4 | 2.9 | 724 | 21 | Aspergillus oryzae |
| 3 | 40.6 | 2.6 | 552 | 21 | Aspergillus oryzae |
| 4 | 39.2 | 2.5 | 1551 | 21 | Arabidopsis thaliana |
| 5 | 39.2 | 2.5 | 1554 | 21 | Arabidopsis thaliana |
| 6 | 39.2 | 2.5 | 1564 | 21 | Arabidopsis thaliana |
| 7 | 39.2 | 2.5 | 5839 | 19 | Hydra head activat |
| 8 | 38.2 | 2.4 | 6741 | 21 | Gene encoding a su |
| 9 | 37.8 | 2.4 | 3510 | 16 | Mitochondrial cyto |

| | | | | | | | |
|---|----|------|-----|-------|----|-----------|----------------------|
| c | 10 | 37.6 | 2.4 | 1686 | 21 | AAC38835 | Arabidopsis thaliana |
| | 11 | 36.6 | 2.3 | 10732 | 21 | AAAL0594 | Gene encoding a su |
| | 12 | 36.6 | 2.3 | 9960 | 21 | AAZ50905 | Human TBC-1 Partia |
| | 13 | 36.2 | 2.3 | 1900 | 12 | AAQ14834 | Cytochrome P450C25 |
| c | 14 | 35.4 | 2.2 | 2036 | 21 | AAF07981 | Fusarium venenatum |
| | 15 | 35 | 2.2 | 4590 | 22 | AAH24065 | Yeast AOD9604-asso |
| | 16 | 34.6 | 2.2 | 446 | 22 | AAAL14748 | Human breast cancer |
| | 17 | 34.6 | 2.2 | 550 | 22 | AAAL23610 | Human breast cancer |
| | 18 | 34.4 | 2.2 | 291 | 21 | AAZ90043 | Hydrophobic domain |
| | 19 | 34.4 | 2.2 | 341 | 21 | AAZ42802 | Human 5' EST Isola |
| | 20 | 34.4 | 2.2 | 677 | 22 | AAH07357 | Human cDNA clone (|
| | 21 | 34.4 | 2.2 | 1045 | 22 | AAH14780 | Human cDNA sequenc |
| | 22 | 34.4 | 2.2 | 1189 | 21 | AAZ90053 | Hydrophobic domain |
| | 23 | 34.4 | 2.2 | 5059 | 20 | AAH84332 | Stealth virus nucl |
| | 24 | 34.2 | 2.2 | 15359 | 22 | ABA08724 | Human ryanodine re |
| | 25 | 34.2 | 2.2 | 15359 | 23 | AA569156 | DNA encoding novel |
| c | 26 | 34 | 2.2 | 627 | 22 | AA198766 | Human excretory re |
| | 27 | 34 | 2.2 | 627 | 22 | AA163162 | Human kidney relat |
| | 28 | 34 | 2.2 | 1473 | 18 | AAAT64548 | NADP-specific glut |
| | 29 | 34 | 2.2 | 1506 | 18 | AAAT64547 | NADP-specific glut |
| c | 30 | 34 | 2.2 | 1969 | 18 | AAAT64531 | Glutamate dehydrog |
| | 31 | 34 | 2.2 | 2096 | 18 | AAAT64543 | NADP-specific glut |
| | 32 | 34 | 2.2 | 2099 | 18 | AAAT64530 | NADP-specific glut |
| | 33 | 34 | 2.2 | 2137 | 18 | AAAT64542 | NADP-specific glut |
| c | 34 | 34 | 2.2 | 2140 | 18 | AAAT64529 | DNA fragment carry |
| | 35 | 34 | 2.2 | 5331 | 21 | AAAS30292 | Drosophila melanog |
| | 36 | 33.8 | 2.1 | 3367 | 23 | ABLO2489 | Drosophila melanog |
| | 37 | 33.8 | 2.1 | 10997 | 23 | ABLO2488 | Human nervous syst |
| | 38 | 33.6 | 2.1 | 2724 | 22 | ABA15116 | Novel human diagno |
| c | 39 | 33.2 | 2.1 | 410 | 22 | AAAS36979 | DNA encoding novel |
| | 40 | 33.2 | 2.1 | 2127 | 23 | AAAS88963 | Drosophila melanog |
| | 41 | 33.2 | 2.1 | 3168 | 23 | ABLI2363 | DNA encoding novel |
| | 42 | 33 | 2.1 | 1038 | 23 | AA573881 | Drosophila melanog |
| c | 43 | 33 | 2.1 | 1109 | 23 | ABLO3511 | Drosophila melanog |
| | 44 | 33 | 2.1 | 1977 | 22 | AAK53275 | Human polynucleoti |
| | 45 | 33 | 2.1 | 1978 | 23 | AAAS73883 | DNA encoding novel |

ALIGNMENTS

RESULT 1

AAO02693

ID AAO02693 standard; DNA; 1764 BP.

AC AAO02693;

XX 02-MAY-2001 (first entry)

XX Exophiala spinifera permease DNA.

XX Fumonisin; degradative enzyme; transport enzyme; flavin monooxygenase;
KW aldehyde dehydrogenase; permease; p-glycoprotein; fumonisin esterase;
KW detoxification; mycotoxin; animal feed; human feed; silage;
KW transgenic plant; transgenic animal; microbial spray; ds.

XX Exophiala spinifera.

XX Key Location/Qualifiers

FT CDS 5..1582

FT /tag= a

FT /product= "E. spinifera permease"

FT /note= "this region is specifically claimed in

FT claim 1b as SEQ ID NO: 7"

XX WO200105980-A1.

XX 25-JAN-2001.

XX 14-JUL-1999; 99WO-US15824.

XX 14-JUL-1999; 99WO-US15824.


```
QY 1561 ttccaagtgggccaatga 1578
Db 1565 ttccaagtgggccaatga 1582

RESULT 2
AAFI2818
ID AAF12818 standard; cDNA: 724 BP.
AC AAF12818;
XX
XX
DT 13-MAR-2001 (first entry)
DE Aspergillus oryzae EST SEQ ID NO:5341.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Aspergillus oryzae.
XX
XX WO200056762-A2.
PN
XX
XX 28-SEP-2000.
XX
XX 22-MAR-2000; 2000WO-US07781.
XX
XX 22-MAR-1999; 99US-0273623.
XX
XX (NOVO ) NOVO NORDISK BIOTECH INC.
XX (NOVO ) NOVO NORDISK AS.
XX
XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX WPI; 2000-594572/56.
XX
XX Monitoring differential expression of genes in filamentous fungal cells
XX uses fluorescence-labeled nucleic acids isolated from the cells and a
XX substrate of expressed sequence tags -
XX
XX Claim 88; Page 2226; 3161pp; English.
XX
XX The present invention describes a method for monitoring differential
XX expression of genes in a first filamentous fungal (FF) cell relative to
XX expression of the same genes in one or more second filamentous fungal
XX cells. The method uses fluorescence-labeled nucleic acids isolated from
XX the FF cells and a substrate of expressed sequence tags (EST). The ESTs
XX are used in the methods for monitoring differential expression of genes
XX in a first filamentous fungal (FF) cell relative to expression of the
XX same genes in one or more second filamentous fungal cells. Monitoring
XX the global expression of genes from FF cells allows the production
XX potential of the microorganisms to be improved. New genes may be
XX discovered, possible functions of unknown open reading frames can be
XX identified and gene copy number variation and stability can be
XX monitored. The expression of genes can be used to study how FF cells
XX adapt to changes in culture conditions, environmental stress, spore
XX morphogenesis, recombination, metabolic or catabolic pathway
XX engineering. Using ESTs provides several advantages over genomic or
XX random cDNA clones including elimination of redundancy as one spot on an
XX array equals one gene or open reading frame, and organisation of the
XX microarrays based on function of the gene products to facilitate
XX analysis of the results. AAF07478 to AAF11247 represents ESTs from
XX Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
XX niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
XX AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
XX all specifically claimed in the present invention.
XX
XX Sequence 724 BP; 153 A; 221 C; 175 G; 175 T; 0 other;
XX
XX
```

```
Query Match 2.9%; Score 46.4; DB 21; Length 724;
Best Local Similarity 46.7%; Pred. No. 0.00031;
Matches 183; Conservative 0; Mismatches 206; Indels 3; Gaps 1;

QY 757 ttcatggcttgagcagcaatgcattcgtgctgaagaatcacacagatgctgctgtacg 816
Db 332 tttagctggcctcgacggagtaatacctacatggccgaagaagtcaaaaacgccaagtgtgc 391
QY 817 gtacccaaagcagtggtcagtgcaatcataattggctctctgcacgcctttccatataca 876
Db 392 gtgcacggggccatgatctctcgcacatgatcaacggaaccccttgcttgcctacctc 451
QY 877 atcgcagttctgttggaattacagatct---cgactctattcttaagtccgcggctat 933
Db 452 atcgcagacctgtaactgcatggcgactacacacggcgctgtagcggcggtac 511
QY 934 attccattcgagacaatgacgcagctcccttcggtcgtcagtttgcaacggctctc 993
Db 512 cgcatcatcacatcgcctaccacagccacgggctccaaagcagcaacctagctgtg 571
QY 994 tgtggcggtatcgtgatggccttcttcgcctcacaacgctgtacaaagagactgctcga 1053
Db 572 gcgatggcgatgctcccggctggatgcgcctcttaacggtctcgcctcgttaaccgc 631
QY 1054 ctcaactggagctttgccgggacaatgggtgggtattttccactcatctcgaaacgca 1113
Db 632 ctcaactggcgatcgcacgcgataacggcctccctcttctcgacttcttcgcgcgcgc 691
QY 1114 catccccgctggcaagtctcctgttggctct 1145
Db 692 gacccaacctataaatacccccctcgcgcctct 723

RESULT 3
AAFI2587
ID AAF12587 standard; cDNA: 552 BP.
XX
XX AAF12587;
XX
XX 13-MAR-2001 (first entry)
XX
XX Aspergillus oryzae EST SEQ ID NO:5110.
XX
XX Multiple gene expression; filamentous fungal cell; EST;
XX expressed sequence tag; Fusarium venenatum; Aspergillus niger;
XX Aspergillus oryzae; Trichoderma reesei; identification; recombination;
XX culture condition; environmental stress; spore morphogenesis;
XX metabolic pathway engineering; catabolic pathway engineering; ss.
XX
XX Aspergillus oryzae.
XX OS
XX WO200056762-A2.
XX
XX 28-SEP-2000.
XX
XX 22-MAR-2000; 2000WO-US07781.
XX
XX 22-MAR-1999; 99US-0273623.
XX
XX (NOVO ) NOVO NORDISK BIOTECH INC.
XX (NOVO ) NOVO NORDISK AS.
XX
XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX WPI; 2000-594572/56.
XX
XX Monitoring differential expression of genes in filamentous fungal cells
XX uses fluorescence-labeled nucleic acids isolated from the cells and a
XX substrate of expressed sequence tags -
XX
XX Claim 88; Page 2142; 3161pp; English.
XX
XX The present invention describes a method for monitoring differential
```



```
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 17-JUL-1999; 99US-0144325.  
PR 18-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145919.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 08-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 15-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.
```

```
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 13-OCT-1999; 99US-0159325.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 23-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
PR 29-OCT-1999; 99US-0162143.
```

Query Match 2.5%; Score 39.2; DB 21; Length 1551;

Best Local Similarity 51.7%; Pred. No. 0.11;

Matches 89; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

```
QY 176 ctggtatctcaggcagctctccagctccctactagcggggggccgctcactctctt 235  
Db 149 ctggtatcaccaccactcaacacgggttaagattcggcggaactgtlcaactctggct 208  
QY 236 acggcatcttaatactactctctctacatctgctacatctgcttcatttcattagccgaactga 295  
Db 209 acggatggttcctcggcggtcctctccacaatgtgggtgggttatctatggcggatct 268  
QY 296 caagcgtctaccgactgcgggtggccaatacatatttgcgtcgtacctggc 347  
Db 269 gctcctctaccctacccctcgggtggtctctactactcggagtgctatgctcgc 320
```

RESULT 5

AAC43230

ID AAC43230 standard; DNA; 1554 BP.

XX AAC43230;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 38505.

XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 9905-0121825.
PR 05-MAR-1999; 9905-0123180.
PR 09-MAR-1999; 9905-0123548.
PR 23-MAR-1999; 9905-0125788.
PR 25-MAR-1999; 9905-0126264.
PR 29-MAR-1999; 9905-0126785.
PR 01-APR-1999; 9905-0127462.
PR 06-APR-1999; 9905-0128234.
PR 08-APR-1999; 9905-0128714.
PR 16-APR-1999; 9905-0129845.
PR 19-APR-1999; 9905-0130077.
PR 21-APR-1999; 9905-0130449.
PR 23-APR-1999; 9905-0130510.
PR 28-APR-1999; 9905-0130891.
PR 30-APR-1999; 9905-0131449.
PR 30-APR-1999; 9905-0132048.
PR 30-APR-1999; 9905-0132407.
PR 04-MAY-1999; 9905-0132484.
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 07-MAY-1999; 9905-0132487.
PR 11-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 17-JUN-1999; 9905-0139453.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144684.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 23-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147036.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151338.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.

| | | | | |
|-----------------|---------------|----|--------------|-----------------|
| PR 28-SEP-1999; | 99US-0156458. | XX | 25-FEB-2000; | 2000EP-0301439. |
| PR 29-SEP-1999; | 99US-0156596. | PP | | |
| PR 04-OCT-1999; | 99US-0157117. | XX | | |
| PR 05-OCT-1999; | 99US-0157753. | PR | 25-FEB-1999; | 99US-0121825. |
| PR 06-OCT-1999; | 99US-0157865. | PR | 05-MAR-1999; | 99US-0123180. |
| PR 07-OCT-1999; | 99US-0158029. | PR | 09-MAR-1999; | 99US-0123548. |
| PR 08-OCT-1999; | 99US-0158232. | PR | 23-MAR-1999; | 99US-0125788. |
| PR 12-OCT-1999; | 99US-0158369. | PR | 25-MAR-1999; | 99US-0126264. |
| PR 13-OCT-1999; | 99US-0158369. | PR | 29-MAR-1999; | 99US-0126785. |
| PR 13-OCT-1999; | 99US-0158294. | PR | 01-APR-1999; | 99US-0127462. |
| PR 13-OCT-1999; | 99US-0158295. | PR | 06-APR-1999; | 99US-0128234. |
| PR 13-OCT-1999; | 99US-0158329. | PR | 08-APR-1999; | 99US-0128714. |
| PR 14-OCT-1999; | 99US-0159330. | PR | 16-APR-1999; | 99US-0129845. |
| PR 14-OCT-1999; | 99US-0159331. | PR | 19-APR-1999; | 99US-0130077. |
| PR 14-OCT-1999; | 99US-0159337. | PR | 21-APR-1999; | 99US-0130449. |
| PR 14-OCT-1999; | 99US-0159638. | PR | 23-APR-1999; | 99US-0130510. |
| PR 18-OCT-1999; | 99US-0159584. | PR | 23-APR-1999; | 99US-0130891. |
| PR 21-OCT-1999; | 99US-0160741. | PR | 28-APR-1999; | 99US-0131449. |
| PR 21-OCT-1999; | 99US-0160767. | PR | 30-APR-1999; | 99US-0132048. |
| PR 21-OCT-1999; | 99US-0160768. | PR | 04-MAY-1999; | 99US-0132407. |
| PR 21-OCT-1999; | 99US-0160770. | PR | 04-MAY-1999; | 99US-0132484. |
| PR 21-OCT-1999; | 99US-0160814. | PR | 05-MAY-1999; | 99US-0132485. |
| PR 21-OCT-1999; | 99US-0160815. | PR | 06-MAY-1999; | 99US-0132486. |
| PR 22-OCT-1999; | 99US-0160980. | PR | 07-MAY-1999; | 99US-0132487. |
| PR 22-OCT-1999; | 99US-0160981. | PR | 11-MAY-1999; | 99US-0132863. |
| PR 25-OCT-1999; | 99US-0161404. | PR | 11-MAY-1999; | 99US-0134256. |
| PR 25-OCT-1999; | 99US-0161405. | PR | 14-MAY-1999; | 99US-0134218. |
| PR 25-OCT-1999; | 99US-0161406. | PR | 14-MAY-1999; | 99US-0134221. |
| PR 26-OCT-1999; | 99US-0161359. | PR | 14-MAY-1999; | 99US-0134370. |
| PR 26-OCT-1999; | 99US-0161360. | PR | 18-MAY-1999; | 99US-0134768. |
| PR 26-OCT-1999; | 99US-0161361. | PR | 18-MAY-1999; | 99US-0134941. |
| PR 28-OCT-1999; | 99US-0161920. | PR | 19-MAY-1999; | 99US-0135124. |
| PR 28-OCT-1999; | 99US-0161992. | PR | 20-MAY-1999; | 99US-0135353. |
| PR 28-OCT-1999; | 99US-0161993. | PR | 21-MAY-1999; | 99US-0135353. |
| PR 29-OCT-1999; | 99US-0162142. | PR | 24-MAY-1999; | 99US-0135629. |
| | | | 25-MAY-1999; | 99US-0136021. |
| | | | 27-MAY-1999; | 99US-0136392. |
| | | | 28-MAY-1999; | 99US-0136782. |
| | | | 01-JUN-1999; | 99US-0137222. |
| | | | 01-JUN-1999; | 99US-0137222. |
| | | | 03-JUN-1999; | 99US-0137528. |
| | | | 04-JUN-1999; | 99US-0137502. |
| | | | 07-JUN-1999; | 99US-0137724. |
| | | | 08-JUN-1999; | 99US-0138094. |
| | | | 10-JUN-1999; | 99US-0138540. |
| | | | 10-JUN-1999; | 99US-0138847. |
| | | | 14-JUN-1999; | 99US-0139119. |
| | | | 16-JUN-1999; | 99US-0139452. |
| | | | 16-JUN-1999; | 99US-0139453. |
| | | | 17-JUN-1999; | 99US-0139492. |
| | | | 18-JUN-1999; | 99US-0139454. |
| | | | 18-JUN-1999; | 99US-0139455. |
| | | | 18-JUN-1999; | 99US-0139456. |
| | | | 18-JUN-1999; | 99US-0139457. |
| | | | 18-JUN-1999; | 99US-0139458. |
| | | | 18-JUN-1999; | 99US-0139459. |
| | | | 18-JUN-1999; | 99US-0139460. |
| | | | 18-JUN-1999; | 99US-0139461. |
| | | | 18-JUN-1999; | 99US-0139462. |
| | | | 18-JUN-1999; | 99US-0139463. |
| | | | 18-JUN-1999; | 99US-0139750. |
| | | | 18-JUN-1999; | 99US-0139763. |
| | | | 21-JUN-1999; | 99US-0139817. |
| | | | 22-JUN-1999; | 99US-0139899. |
| | | | 23-JUN-1999; | 99US-0140353. |
| | | | 23-JUN-1999; | 99US-0140354. |
| | | | 24-JUN-1999; | 99US-0140695. |
| | | | 28-JUN-1999; | 99US-0140823. |
| | | | 29-JUN-1999; | 99US-0140991. |
| | | | 30-JUN-1999; | 99US-0141287. |
| | | | 01-JUL-1999; | 99US-0141842. |
| | | | 01-JUL-1999; | 99US-0142154. |
| | | | 02-JUL-1999; | 99US-0142055. |
| | | | 06-JUL-1999; | 99US-0142390. |

Query Match 2.5%; Score 39.2; DB 21; Length 1554;
Best Local Similarity 51.7%; Pred. No. 0.11; Mismatches 83; Indels 0; Gaps 0;
Matches 89; Conservative 0;

| | |
|--------|--|
| Qy 176 | ctggatctcaggcagctctccagctcgcctactagcggggggccgcctactctctt 235 |
| Db | |
| Qy 149 | ctggatctcaccacactacacacacggcttaagattcgccgactgtcactgtgtct 208 |
| Db | |
| Qy 236 | acggatctcactaactcagctactctctctacactctgctcttctcattagccgaactga 295 |
| Db | |
| Qy 296 | ccagctctaccgactgcggctggccaatatcatttgcgtcgatcctggc 347 |
| Db | |
| Qy 269 | gctcctctaccctactcctcgggtgtctctactactgagtgatgctgcgc 320 |

RESULT 6
AAC49990
ID AAC49990 standard; DNA; 1664 BP.
XX AC AAC49990;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 63189.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.

```
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.

PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157253.
PR 08-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159333.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.5%; Score 39.2; DB 21; Length 1664;
Best Local Similarity 51.7%; Pred. No. 0.11;
Matches 89; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

Qy 176 ctggtatctcaaggcagctccagctcgccctactagcggggggggccgctactctctt 235
Db 184 ctggtatctcaaggcagctccagctcgccctactagcggggggggccgctactctctt 243
Qy 236 acggcatcttaatactcagctactctctctacatctgcctcttctcttctatagccgaactga 295
Db 244 acggatggttcctcgccgctctctcaaatgctgcttgggttatctatgcccagatct 303
Qy 296 ccaggctctaccgactcgccgctggccaatatcttttgcgtcgatctctgc 347
Db 304 gctcctctaccctactccggtggtctctactactgagtgctatgctcgc 355

RESULT 7
AAV46335/C
ID AAV46335 standard; cDNA: 5839 BP.
XX
AC AAV46335;
XX
DT 20-NOV-1998 (first entry)
XX
DE Hydra head activator binding protein cDNA.
XX
KW Head activator binding protein; hydra; screening assay; agonist;
XX antagonist; treatment; neurodegenerative disease; ss.
OS Chlorohydra viridissima.
XX
FH Key
FT CDS
FT 481..5466
FT /*tag= a
```

```
FT sig_peptide 481..567
FT /tag= b
FT mat_peptide 568..5463
FT /tag= c
FT /product= "head activator binding protein"
XX DE19808258-A1.
XX 03-SEP-1998.
XX 27-FEB-1998; 98DE-1008258.
XX 10-JUL-1997; 97DE-1029490.
XX 28-FEB-1997; 97DE-1008134.
XX (EVOT-) EVOTEC BIOSYSTEMS GMBH.
XX Hampe W, Schaller C;
XX WPI: 1998-468562/41.
XX P-PSDB: AA476041.
XX Hydra head activator binding protein - useful in screening assay for
XX agonists or antagonists
XX Claim 18; Page 11-14; 21pp; German.
XX This sequence encodes a head activator binding protein isolated from
XX Chlorohydra viridissima by chromatographic methods. The protein is used
XX in a screening assay for agonists or antagonists comprising determining
XX the binding properties of candidate substances with the protein. The
XX assay can be used to screen for substances useful for treating
XX neurodegenerative diseases.
XX Sequence 5839 BP; 1856 A; 978 C; 1106 G; 1899 T; 0 other;
XX
XX Query Match 2.5%; Score 39.2; DB 19; Length 5839;
XX Best Local Similarity 56.18; Pred. No. 0.23;
XX Matches 74; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 1365 ggtggtttacagtcggtcaccactgtgttttcagcttcacactgacgctgctacgacgc 1424
DB 3883 GTTTTCATCATCACCACATCCCCACATCATTTTCACCATCATCAATGCCATGACAAACTTG 3824
QY 1425 cgcgtcaaccatgaattacacagtcgcatattatagcgttgcacttgctgtgtgtctt 1484
DB 3823 GTAAACAACGATTATTTCACATGTGAATGTTAGCGTTGCACCTTTACTTTATTGCATA 3764
QY 1485 gaactgggtcgt 1496
DB 3763 TATTTTGTCTT 3752
XX
RESULT 8
ID AAA10595/c
XX AAA10595 standard; DNA; 6741 BP.
XX AAA10595;
XX
XX 29-JUN-2000 (first entry)
XX
XX Gene encoding a subunit of cellulose synthase.
XX Cellulose synthase; cellulose production; increase yield; ds.
XX Vigna angularis.
XX JP2000060568-A.
XX
XX 29-FEB-2000.
XX
XX 26-AUG-1998; 98JP-0239998.
XX
XX 26-AUG-1998; 98JP-0239998.
XX (MIZU/) MIZUNO K.
XX (OJIP ) OJI PAPER CO.
XX WPI: 2000-342371/30.
XX P-PSDB: AA485180.
XX
XX A gene encoding a cellulose synthetic equipment - for the improvement
XX in the amount of cellulose synthesised in a plant body
XX Claim 2; Page 22-31; 32pp; Japanese.
XX
XX This sequence represents a gene encoding a subunit of the cellulose
XX synthase complex of Vigna angularis. The invention relates to subunits of
XX cellulose synthetic equipment, that can be used to increase the amount of
XX cellulose synthesised by a plant. The proteins and genes encoding them
XX can also be used to improve the properties of the cellulose being
XX produced by a plant.
XX Sequence 6741 BP; 1712 A; 870 C; 1468 G; 1374 T; 1317 other;
XX
XX Query Match 2.4%; Score 38.2; DB 21; Length 6741;
XX Best Local Similarity 19.6%; Pred. No. 0.52;
XX Matches 46; Conservative 88; Mismatches 101; Indels 0; Gaps 0;
QY 126 tagcaccatcaccatcgtctccttgcccttgattgcaacagttggcgtggtatctc 185
DB 5880 TDRCYRACRSYNCRCYSYSSYDATYTBRSRGYSNTRCDYDAYSDYDASTSDTTS 5821
QY 186 agcgagtcctccagtcgcctactagcggggggggcgcgtcactctccttaccgcatcct 245
DB 5820 RSDSTTCTYTNSTNSRDNSTCTRCYSSRCYTYSYSTCTSDYSYRAAKCYTYSRST 5761
QY 246 aatcagtcactctcgtacatcgtcgtcttctcattagccggaactgaccagcgtcta 305
DB 5760 NSTYYDAONSTTBSTTCYTTTNTNCYNSTYSYRANSTTBVRCYRACYSCYCYT 5701
QY 306 ccgactcgcgggtggcgaatacattttgcgcgactcgtgcgacccaatcaatc 360
DB 5700 YSDASTCYTSRGTCYTTBYSTFTBNCRCNCYSDSYRANCYSSTYDABNSTY 5646
XX
RESULT 9
ID AAQ92829 standard; DNA; 3510 BP.
XX AAQ92829;
XX
XX 21-FEB-1996 (first entry)
XX
XX Mitochondrial cytochrome P450/NADPH P450 reductase fusion DNA.
XX Mitochondrial; cytochrome P450; NADPH; reductase; mature c25;
XX Industrial waste; biological oxidation; fusion DNA; ds.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 1..3510
XX /*tag= a
XX
XX JP07147975-A.
XX
XX 13-JUN-1995.
XX
XX 29-NOV-1993; 93JP-0298279.
XX
XX 29-NOV-1993; 93JP-0298279.
XX
XX (SUMO ) SUNITOMO CHEM CO LTD.
```



```
FT intron 93591..97475
FT /*tag= x
FT /label= K
FT 97476..97960
FT /*tag= y
FT /label= L
FT 988..1006
FT /*tag= z
FT /bound_moiety= "Primer B2"
FT /note= "Amplification of amplicon 99-20508"
FT complement (1509..1529)
FT /*tag= aa
FT /bound_moiety= "Primer C2"
FT /note= "Amplification of amplicon 99-20508"
FT 5039..5056
FT /*tag= ab
FT /bound_moiety= "Primer B3"
FT /note= "Amplification of amplicon 99-20469"
FT complement (5534..5554)
FT /*tag= ac
FT /bound_moiety= "Primer C3"
FT /note= "Amplification of amplicon 99-20469"
FT 5997..6015
FT /*tag= ad
FT /bound_moiety= "Primer B4"
FT /note= "Amplification of amplicon 5-254"
FT complement (6332..6350)
FT /*tag= ae
FT /bound_moiety= "Primer C4"
FT /note= "Amplification of amplicon 5-254"
FT 14371..14390
FT /*tag= af
FT /bound_moiety= "Primer B5"
FT /note= "Amplification of amplicon 5-257"
FT complement (14798..14817)
FT /*tag= ag
FT /bound_moiety= "Primer C5"
FT /note= "Amplification of amplicon 5-257"
FT 18751..18771
FT /*tag= ah
FT /bound_moiety= "Primer B6"
FT /note= "Amplification of amplicon 99-20511"
FT complement (19198..19217)
FT /*tag= ai
FT /bound_moiety= "Primer C6"
FT /note= "Amplification of amplicon 99-20511"
FT 19605..19625
FT /*tag= aj
FT /bound_moiety= "Primer B7"
FT /note= "Amplification of amplicon 99-20510"
FT complement (19986..20005)
FT /*tag= ak
FT /bound_moiety= "Primer C7"
FT /note= "Amplification of amplicon 99-20510"
FT 29529..29547
FT /*tag= al
FT /bound_moiety= "Primer B8"
FT /note= "Amplification of amplicon 99-20504"
FT complement (30041..30061)
FT /*tag= am
FT /bound_moiety= "Primer C8"
FT /note= "Amplification of amplicon 99-20504"
FT 42268..42287
FT /*tag= an
FT /bound_moiety= "Primer B9"
FT /note= "Amplification of amplicon 99-20493"
FT complement (42732..42752)
FT /*tag= ao
FT /bound_moiety= "Primer C9"
FT /note= "Amplification of amplicon 99-20493"
FT 69026..69046
FT /*tag= ap
FT /bound_moiety= "Primer B10"
```

```
FT primer_bind
FT /note= "Amplification of amplicon 99-20499"
FT complement (69525..69543)
FT /*tag= aq
FT /bound_moiety= "Primer C10"
FT /note= "Amplification of amplicon 99-20499"
FT 76323..76343
FT /*tag= ar
FT /bound_moiety= "Primer B11"
FT /note= "Amplification of amplicon 99-20473"
FT complement (76771..76790)
FT /*tag= as
FT /bound_moiety= "Primer C11"
FT /note= "Amplification of amplicon 99-20473"
FT 78292..78309
FT /*tag= at
FT /bound_moiety= "Primer B12"
FT /note= "Amplification of amplicon 5-249"
FT complement (78704..78721)
FT /*tag= au
FT /bound_moiety= "Primer C12"
FT /note= "Amplification of amplicon 5-249"
FT 81893..81912
FT /*tag= av
FT /bound_moiety= "Primer B13"
FT /note= "Amplification of amplicon 99-20485"
FT complement (82353..82372)
FT /*tag= aw
FT /bound_moiety= "Primer C13"
FT /note= "Amplification of amplicon 99-20485"
FT 84392..84412
FT /*tag= ax
FT /bound_moiety= "Primer B14"
FT /note= "Amplification of amplicon 99-20481"
FT complement (84909..84929)
FT /*tag= ay
FT /bound_moiety= "Primer C14"
FT /note= "Amplification of amplicon 99-20481"
FT 89746..89765
FT /*tag= az
FT /bound_moiety= "Primer B15"
FT /note= "Amplification of amplicon 99-20480"
FT complement (90179..90198)
FT /*tag= ba
FT /bound_moiety= "Primer C15"
FT /note= "Amplification of amplicon 99-20480"
FT 1431..1455
FT /*tag= bb
FT /bound_moiety= "Probe P2"
```

```
Query Match 2.3%; Score 36.6; DB 21; Length 99960;
Best Local Similarity 54.0%; Pred. No. 8.4;
Matches 75; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
Qy 359 tcaatcggagcatttcacgtgtgcgactcgtgtcgttcttcattgcgtatcgc 418
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24794 tgaacggagdgctgcctgtgttcttgcactgccctcgtgaagtgtgtggaagcgtggg 24853
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 419 gaagctcagtgaccatgacatcgtgtcacaagatccgcgcgtgatagccgcctatgctc 478
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24854 gaagccaagtcctcattggtgcctccatcaggagacctgcagctggagcagcagagg 24913
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 479 acacatactcccaggattc 497
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24914 ccacagggtgtgtagcttc 24932
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 13
AAQ14834
ID AAQ14834 standard; DNA: 1900 BP.
XX
XX AAQ14834;
XX
DT 10-FEB-1992 (first entry)
```

```
XX DE Cytochrome P450C25 gene.
XX KW Vitamin D3; ss.
XX OS Rattus rattus.
XX FH Location/Qualifiers
XX FT 59..1660
XX FT /*tag= a
XX FT sig_peptide 59..154
XX FT /*tag= b
XX FT mat_peptide 155..1657
XX FT /*tag= c
XX PN JP03232493-A.
XX PD 16-OCT-1991.
XX PF 06-FEB-1990; 90JP-0027711.
XX PR 06-FEB-1990; 90JP-0027711.
XX PA (SUMO ) SUMITOMO CHEM IND KK.
XX DR WPI; 1991-349005/48.
XX DR P-PSDB; AAR15057.
XX CY Cytochrome P450C25 gene expressed by microbe cells - is used to
XX PT produce enzyme for mfr. of active vitamin-D3, etc.
XX PS Claim 1; Fig 1; 10pp; Japanese.
XX CC The sequence is that of the cytochrome P450C25 gene which was obt'd.
XX CC from rat liver. It can be recombinantly expressed in microorganisms
XX CC in sufficient amounts to be used in the process of manufacturing
XX CC active vitamin D3.
XX SQ Sequence 1900 BP; 451 A; 552 C; 476 G; 421 T; 0 other;

Query Match 2.3%; Score 36.2; DB 12; Length 1900;
Best Local Similarity 59.0%; Pred. No. 1.1;
Matches 62; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 6 ctcacagaccagtgatcacggcgagaaaggcgagacagagcagacagacagacagacaga 65
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 66 gacggcgggcgagtggtggtgcgtccgagtcctgaaacttctct 110
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 222 gctcgcggagtgctgagcgagcttcgcgggaccgggacgcgtacgct 266

RESULT 14
AAF07981
ID AAF07981 standard; cDNA; 2036 BP.
XX AC AAF07981;
XX 13-MAR-2001 (first entry)
XX DE Fusarium venenatum EST SEQ ID NO:504.
XX KW Multiple gene expression; filamentous fungal cell; EST;
XX KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
XX KW Aspergillus oryzae; Trichoderma reesei; Identification; recombination;
XX KW culture condition; environmental stress; spore morphogenesis;
XX KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX OS Fusarium venenatum.
XX PN WO2000056762-A2.
```

```
XX PD 28-SEP-2000.
XX 22-MAR-2000; 2000WO-US07781.
XX 22-MAR-1999; 99US-0273623.
XX (NOVO ) NOVO NORDISK BIOTECH INC.
XX (NOVO ) NOVO NORDISK AS.
XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX WPI; 2000-594572/56.
XX Monitoring differential expression of genes in filamentous fungal cells
XX PT uses fluorescence-labeled nucleic acids isolated from the cells and a
XX PT substrate of expressed sequence tags -
XX PS Claim 86; Page 578; 3161pp; English.
XX CC The present invention describes a method for monitoring differential
XX CC expression of genes in a first filamentous fungal (FF) cell relative to
XX CC expression of the same genes in one or more second filamentous fungal
XX CC cells. The method uses fluorescence-labeled nucleic acids isolated from
XX CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
XX CC are used in the methods for monitoring differential expression of genes
XX CC in a first filamentous fungal (FF) cell relative to expression of the
XX CC same genes in one or more second filamentous fungal cells. Monitoring
XX CC the global expression of genes from FF cells allows the production
XX CC potential of the microorganisms to be improved. New genes may be
XX CC discovered, possible functions of unknown open reading frames can be
XX CC identified and gene copy number variation and stability can be
XX CC monitored. The expression of genes can be used to study how FF cells
XX CC adapt to changes in culture conditions, metabolic or catabolic pathway
XX CC morphogenesis, recombination, environmental stress, spore
XX CC random cDNA clones including elimination of redundancy as one spot on an
XX CC array equals one gene or open reading frame, and organization of the
XX CC microarrays based on function of the gene products to facilitate
XX CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
XX CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
XX CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
XX CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
XX CC all specifically claimed in the present invention.
XX SQ Sequence 2036 BP; 399 A; 582 C; 523 G; 521 T; 11 other;

Query Match 2.2%; Score 35.4; DB 21; Length 2036;
Best Local Similarity 47.5%; Pred. No. 2.1;
Matches 105; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 130 accatcacatcgctgctcctggcctttgtgattgcaacagttgggtgatctatcaggc 189
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 190 agtctcagctcgccctactagcggggggcccgctactctctcttaccgcatctaatc 249
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 296 accttggttaccctctgattgagcggggccctgtcaacgcttatctggggatggctcgcc 355
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 250 agtactctgctacatctgcacgcctttctctattagcgaactgaccagcgtctaccgc 309
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 356 gtctctctcattattgtgtgctgctgtctctctgttggaatcacacagtggttatcct 415
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 310 actgcggggtggccaatcatcttttcgctgcgtacccggcacc 350
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 416 acaactggaagtgtctactacccaagctttcatgctctcccc 456
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
AAH24065/C
ID AAH24065 standard; DNA; 4590 BP.
XX
```

AC AAR24065;
XX 29-AUG-2001 (first entry)
XX Yeast AOD9604-associated DNA sequence, SEQ ID NO:1.
XX Human growth hormone analogue peptide; hGH; AOD9604; lipid metabolism;
KW modulation; lipolysis stimulation; hormone-sensitive lipase stimulation;
KW lipogenesis inhibition; acetyl CoA carboxylase inhibition; obesity;
XX functional food; transgenic yeast; fat/lean ratio; food use; ds.
OS Saccharomyces cerevisiae.
XX key Location/Qualifiers
FH misc_feature 10
FT /tag= a
FT /note= "Represented as * in the specification"
FT 3617
FT /tag= b
FT /note= "Represented as * in the specification"
FT 3649
FT /tag= c
FT /note= "Represented as * in the specification"
FT 3679
FT /tag= d
FT /note= "Represented as * in the specification"
FT 3819
FT /tag= e
FT /note= "Represented as * in the specification"
FT 3862
FT /tag= f
FT /note= "Represented as * in the specification"
FT 3864
FT /tag= g
FT /note= "Represented as * in the specification"
FT 3888
FT /tag= h
FT /note= "Represented as * in the specification"
FT 3900
FT /tag= i
FT /note= "Represented as * in the specification"
FT 3912
FT /tag= j
FT /note= "Represented as * in the specification"
FT 3914
FT /tag= k
FT /note= "Represented as * in the specification"
FT 3938
FT /tag= l
FT /note= "Represented as * in the specification"
FT 3939
FT /tag= m
FT /note= "Represented as * in the specification"
FT 3941
FT /tag= o
FT /note= "Represented as * in the specification"
FT 3943
FT /tag= p
FT /note= "Represented as * in the specification"
FT 4361
FT /tag= q
FT /note= "Represented as * in the specification"
XX WO200133977-A1.
XX 17-MAY-2001.
XX 06-NOV-2000; 2000WO-AU01362.
XX 05-NOV-1999; 99AU-0003875.
XX (META-) METABOLIC PHARM LTD.

PI Belyea CI, Ng FM, Vaughan P;
XX WPI; 2001-328876/34.
XX New organisms containing nucleic acid encoding a growth hormone
PT fragment which modulates lipid metabolism are useful to produce dietary
PT aids for obesity and in the meat production industry -
XX Disclosure; Page 48-50; 54pp; English.
XX The invention relates to novel transgenic organisms useful in the
CC production of functional food and drink products for the treatment
CC or prevention of obesity via the regulation of lipid metabolism. The
CC organisms comprise a polynucleotide encoding a growth hormone fragment
CC capable of stimulating the activity of hormone-sensitive lipase (the key
CC enzyme in lipolysis) and inhibiting acetyl CoA carboxylase (the key
CC enzyme in lipogenesis). The growth hormone fragment preferably contains
CC at least the disulphide-bonded loop of a mammalian growth hormone (but is
CC not the full-length growth hormone) and is optionally linked to an
CC epitope tag or heterologous fusion protein partner. The transgenic
CC organism may be a microorganism used to produce a fermented product
CC (e.g., yeast), or an edible plant or animal or cell thereof. Food or
CC drink made using methods of the invention are used to modify fat/lean
CC ratio, lipid metabolism or food use in a mammal. In particular, the food
CC or drink products may be used to treat or prevent obesity, particularly
CC in humans, and may also be used to improve the fat/lean ratio of
CC livestock raised for meat production. In the exemplification of the
CC invention, the human growth hormone (hGH) fragment analogue AOD9604 was
CC expressed in yeast, optionally fused to the FLAG epitope (AAB73625).
CC The present sequence is described as a DNA sequence from yeast in
CC the sequence listing, but is not further referred to in the
XX specification.
SQ Sequence 4590 BP; 661 A; 384 C; 127 G; 522 T; 2896 other;
Query Match 2.2% Score 35; DB 22; Length 4590;
Best Local Similarity 14.3% Pred. NO. 4, 6;
Matches 38; Conservative 109; Mismatches 119; Indels 0; Gaps 0;
QY 885 tctgtatggaattacagatctcgactctcttcttaagttccgcgcggtatctccattcga 944
DB 861 HNTKSYDGRKSYASDGDWYHNKSNWYHKTHTNTGNSNSYSYAKSHYSANTAKYTD 802
QY 945 gacaatgacgcagtccttcgctgcagtttgcacacggtctctctcatgtggcggtat 1004
DB 801 GDWSTYSHKYATNNSVYBHRKTSKGNCSNTYAGTAMGNMTDANGKSHKTDGSKASRS 742
QY 1005 cgtgatgacctcttcgctccctcaacgctgtacaagacactgcgtctcgaactcactgag 1064
DB 741 SBYNTYDHTYNTYANGDASSTSHBATYNCYNYRYTHYDYSKNSNAGNANGCHNTYAK 682
QY 1065 ctttgcggggacaatgggtctgtatttccactcatctcgaacgcatcattcccccgtg 1124
DB 681 BYKAVTYTDGWSATCNGCRMGTAHSHSTGCTKTHKYSYNNAGYAVTYTDGWSATCNG 622
QY 1125 gcaagtctcgtgttggtctctattcg 1150
DB 621 WYASSHKYHNTKDGSHKTKNTNTNS 596
Search completed: April 27, 2002, 00:42:14
Job time: 18968 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2002, 23:04:52 ; Search time 7811.87 Seconds
(without alignments)
4227.166 Million cell updates/sec

Title: US-09-882-694A-7

Perfect score: 1578

Sequence: 1 atggactccagaccagtg.....aattccaagtggcccatga 1578

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sta.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|----|-------------|
| ----- | | | | | | |

| | | | | | | |
|----|--------|------|--------|----|-------------|--------------|
| 1 | 1555.6 | 98.6 | 1578 | 6 | AX076848 | Sequence |
| 2 | 1555.6 | 98.6 | 1764 | 6 | AX076847 | Sequence |
| 3 | 66.4 | 4.2 | 2475 | 8 | SCYGL077C | S.cerevisia |
| 4 | 66.4 | 4.2 | 2725 | 8 | YSCCVR | Saccharomyc |
| 5 | 58 | 3.7 | 7218 | 6 | I66494 | Sequence 14 |
| 6 | 52.4 | 3.3 | 3724 | 8 | YSCSEQ | Saccharomyc |
| 7 | 47.4 | 3.0 | 125020 | 9 | AF429315 | Homo sapi |
| 8 | 44.8 | 2.8 | 27893 | 8 | SPAC1039 | S.pombe c |
| 9 | 41.2 | 2.6 | 147203 | 8 | AF003687 | Oryza sat |
| 10 | 41 | 2.6 | 125020 | 9 | AF429315 | Homo sapi |
| 11 | 40 | 2.5 | 1996 | 8 | AF053231 | Neurospor |
| 12 | 39.2 | 2.5 | 720 | 8 | CNS01D5D | Botrytis |
| 13 | 39.2 | 2.5 | 5359 | 3 | AF092920 | Chlorohyd |
| 14 | 39.2 | 2.5 | 84957 | 8 | AC006200 | Arabidops |
| 15 | 38.2 | 2.4 | 6741 | 6 | E32987 | Gene encodi |
| 16 | 38.2 | 2.4 | 286550 | 1 | SME591785 | Sinorhizo |
| 17 | 37.8 | 2.4 | 1201 | 10 | RNCYF27S1 | Rattus norv |
| 18 | 37.8 | 2.4 | 1877 | 10 | RATMTCPYFCH | Rat mitochon |
| 19 | 37.8 | 2.4 | 1900 | 6 | E03144 | cDNA encodi |
| 20 | 37.8 | 2.4 | 1900 | 10 | RNCYP27 | Rattus norv |
| 21 | 37.8 | 2.4 | 2300 | 10 | RATP450C | Sprague-Daw |
| 22 | 37.8 | 2.4 | 3510 | 23 | E09419 | Synthetic n |
| 23 | 37.6 | 2.4 | 1726 | 8 | AF019637 | Arabidops |
| 24 | 37.4 | 2.4 | 981 | 11 | CNS063TH1 | Arabidops |
| 25 | 37.4 | 2.4 | 110000 | 2 | AC013622_0 | T3 end of |
| 26 | 37.4 | 2.4 | 110000 | 2 | AC021632_0 | Mus muscu |
| 27 | 37.2 | 2.4 | 1246 | 6 | AX164174 | Mus muscu |
| 28 | 37.2 | 2.4 | 1891 | 8 | SCUGA4 | Sequence |
| 29 | 37.2 | 2.4 | 5794 | 8 | SCYDL210W | S.cerevisia |
| 30 | 37.2 | 2.4 | 39090 | 8 | CS39KBCIV | S.cerevisia |
| 31 | 36.8 | 2.3 | 12062 | 1 | SCD17 | Streptomy |
| 32 | 36.8 | 2.3 | 329709 | 1 | AP002997 | Mesorhizo |
| 33 | 36.6 | 2.3 | 10732 | 6 | E32986 | Gene encodi |
| 34 | 36.6 | 2.3 | 154344 | 9 | AE002028 | Homo sapi |
| 35 | 36.4 | 2.3 | 11514 | 1 | AC0073308 | Deinococc |
| 36 | 36.4 | 2.3 | 69634 | 2 | AC101748 | Mus muscu |
| 37 | 36.4 | 2.3 | 105213 | 2 | AL158217 | Mus muscu |
| 38 | 36.4 | 2.3 | 203188 | 2 | AC104323 | Homo sapi |
| 39 | 36 | 2.3 | 183876 | 9 | AC006390 | Homo sapi |
| 40 | 35.8 | 2.3 | 289 | 6 | AR162089 | Sequence |
| 41 | 35.8 | 2.3 | 289 | 6 | AR166614 | Sequence |
| 42 | 35.8 | 2.3 | 13848 | 1 | AE005104 | Sequence |
| 43 | 35.8 | 2.3 | 62299 | 2 | AC100224 | Halobacte |
| 44 | 35.8 | 2.3 | 145911 | 1 | AF003014 | Mus muscu |
| 45 | 35.8 | 2.3 | 166335 | 2 | AC073799 | Mesorhizo |
| | | | | | | Mus muscu |

ALIGNMENTS

| | | | | | | |
|------------|------------|--|---------------|---------|-----|------------------------|
| RESULT | 1 | | | | | |
| AX076848 | | | | | | |
| LOCUS | AX076848 | Sequence | 7 from Patent | 1578 bp | DNA | linear PAT 22-FEB-2001 |
| DEFINITION | AX076848 | Sequence | 7 from Patent | 1578 bp | DNA | linear PAT 22-FEB-2001 |
| ACCESSION | AX076848 | Sequence | 7 from Patent | 1578 bp | DNA | linear PAT 22-FEB-2001 |
| VERSION | AX076848.1 | GI:13121521 | | | | |
| KEYWORDS | | Exophiala spinifera. | | | | |
| SOURCE | | Exophiala spinifera. | | | | |
| ORGANISM | | Exophiala spinifera | | | | |
| REFERENCE | | 1 (bases 1 to 1578) | | | | |
| AUTHORS | | Duvalck J.P., Maddox J., Gilliam J., Folkerts, O. and Crasta, O.R. | | | | |
| TITLE | | Compositions and methods for fumonisin detoxification | | | | |
| JOURNAL | | Patent: WO 0105980-A 7 25-JAN-2001; | | | | |
| FEATURES | | Pioneer HI-Bred International Inc. (US) ; Curagen Corporation (US) | | | | |
| | | Location/Qualifiers | | | | |
| | | 1..1578 | | | | |
| | | /organism="Exophiala spinifera" | | | | |
| | | /db_xref="taxon:91928" | | | | |
| | | /note="permease, coding sequence" | | | | |
| | | 324 a 446 c 385 g 423 t | | | | |
| BASE COUNT | | | | | | |

| | | | | | | | | | |
|--|-----|---|------|--|--|--|--|--|--|
| ORIGIN | | | | | | | | | |
| Query Match 98.6%; Score 1555.6; DB 6; Length 1578; Best Local Similarity 99.1%; Pred. No. 0; Matches 1564; Conservative 0; Mismatches 14; Indels 0; Gaps 0; | | | | | | | | | |
| QY | 1 | atggactccagacaaagtgatagcgagaaagcgaggaacagacagacaaagaac | 60 | | | | | | |
| Db | 1 | ATGGACTCCAGACCAAGTGGATACGGCTTGAAGGGGGAGGACAGACAAAGAAAC | 60 | | | | | | |
| QY | 61 | acagagagcgggcgagtggtgctgcgagtcctcgaacgcttccctctggaagaaga | 120 | | | | | | |
| Db | 61 | ACAGAGACGGCGGCGAGTGTGCTCCGAGTCCCTGAACGTTCCCTCGAGAAAGAA | 120 | | | | | | |
| QY | 121 | caatttggacacacacacgctgtcttggccttggatttgattgcaacagttggcctggt | 180 | | | | | | |
| Db | 121 | CAATTGGACACATACCATCTGTCTTGGCCTTTTGATTTGCAAGTTGGCTGGT | 180 | | | | | | |
| QY | 181 | atctcagggagtcctcagctgcctactatagcggggggcccgtcactctccttaagcg | 240 | | | | | | |
| Db | 181 | ATCTCAGGCAAGTCTCCAGCTCGCCCTACTAGCGGGGGGCGCGTCACTCTCCTTACGGC | 240 | | | | | | |
| QY | 241 | atctaatcagtagctctcactatcgatcgcttcttctcattagccgaactacacgc | 300 | | | | | | |
| Db | 241 | ATCCTAATCAGTACTCTGCTCTACATCTGCATCGCTTTCTCATTAGCCGAACACTGACCAGC | 300 | | | | | | |
| QY | 301 | gtctaccgactgcgggtggccaataatcttggctgcgactcctggcaccacaataatc | 360 | | | | | | |
| Db | 301 | GTCTACCAGCACTCGCGGTGGCCAATATCATTTTGGCTGCGATCTCTGGCACCAAAATCAATC | 360 | | | | | | |
| QY | 361 | aactggagcaattcacaogtggcggaactcgtctgttcttctcattagcgcgctatcgga | 420 | | | | | | |
| Db | 361 | AATCGGAGCAATTCATACGTGTGCGGACATCGTGTGCTTCTCATTAGCCGAACACTGACCGA | 420 | | | | | | |
| QY | 421 | agctcagtagccatgatactcgtctcgaagatcccgggcgctgtagccgcctatagtcac | 480 | | | | | | |
| Db | 421 | AGCTCAGTAGCCATGATATACCTGCTCAACAGATCCCGCGCGCTGATAGCCGCTATAGTCAC | 480 | | | | | | |
| QY | 481 | acatactccagagatcgtggcgatgcttctcctcactctacgagggagtcgcgctggtg | 540 | | | | | | |
| Db | 481 | ACATACTCCAGGATTCGTGGCATGCTCTTCTTCATCTACGAGGGAGTCGCGCTGGTGTG | 540 | | | | | | |
| QY | 541 | ctctgttcaactgtttccctcgaagaagaaccccttgggttcatagaatcggattcggc | 600 | | | | | | |
| Db | 541 | CTCTGTTCACCTGTGTTGGCTTGAAGAAACCCCTTGGGTTTCATGAAATCGGATTCGGC | 600 | | | | | | |
| QY | 601 | ctcacgactcctctcgtgatctccttatacgccaattctagcgcggtcccaaccccacg | 660 | | | | | | |
| Db | 601 | CTCACGATCGCTCTCTCGTGATCTCCTTTATCGCCATTCCTAGCCGCGTCCCAACCCCAAG | 660 | | | | | | |
| QY | 661 | gctccaaactcacagtagtagctgtagagcaactatactggctggttcgagcgagtc | 720 | | | | | | |
| Db | 661 | GCTCCAAACTCACAGTAGTAGCTGTTGGAGCAACTATACGTGGTGGTTCGACGGCGTC | 720 | | | | | | |
| QY | 721 | tgtctcatcctggccttcgcacactcctccttcattgttcattggttggagcagcaatg | 780 | | | | | | |
| Db | 721 | TGCTTCATCTCGGCCCTTCGACATCTCTGCTCATGTTTCATTGGCTTGGACGACGAATG | 780 | | | | | | |
| QY | 781 | catctggctgaagaatgcacagatgctcgtacggttaccacaagaagatggtcagtcga | 840 | | | | | | |
| Db | 781 | CATCTGGCTGAAGAATGCACAGATGCTGCTGACGGTACCCCAAGCAGTGGTCAGTGCA | 840 | | | | | | |
| QY | 841 | atcataattggctcttgcacccgcttccataatacaactcgagttctgtatggaattaca | 900 | | | | | | |
| Db | 841 | ATCATAATTGGCTCTTCGACCGGCTTTTCCATATACAAATCGCATGTTCTGTATGGAATPACA | 900 | | | | | | |
| QY | 901 | gactcgaactcattctgaagttccggcgctatatctcattcagacagacaacagcagtc | 960 | | | | | | |
| Db | 901 | GACTCGACACTATTCTAGTTCCCGCGGCTATATTCATTCCAGAACATGAGGCGAGTCT | 960 | | | | | | |
| QY | 961 | cttcggtgcgtcagtttggcaacggtctctctcatgtggcggtatcgtgattggcctcttc | 1020 | | | | | | |
| Db | 961 | CTTCGGTGCCTCAGTTTGGCAACGGTCTCTCTCATGTGGCGGTATCGTGATGGCCTTCTTC | 1020 | | | | | | |

QY 1 atgactccagacaaagtggatagcggcgagaaaggcggaagggagacaaagcagacaaagaaac 60
DB 5 ATGACTCCAGACCAAGTGGATACGGCTTGAAGGGCGGACAAAGCGACACAAAGAAC 64
QY 61 acagagacggcgggcgagtggtgctgcagagtcctgaacgttccctctgagaaagaaa 120
DB 65 ACAGAGACGGCGGCGAGGTGCTCGAGTCCCTGAACGTTCTCTGGAGAAGAAA 124
QY 121 caatttgccaccatcaacatcgctgctcgttgcttggtgcttgatgtaaacagattggctggt 180
DB 125 CAATTTGGCGACATCAACATCGTGTCTTGGCCCTTGTGATTTGCAACAGTTGGGCTGGT 184
QY 181 atcacaagcagctccagctcgccctactagcggggggggcccgctcacctctcttaccggc 240
DB 185 ATCTCAGGACGTCPCAGCTCGCCCTACTAGCGGGGGGGCCCGTCACTCTCTCTTACGGC 244
QY 241 atcctaactcagctactcgtctacatcctgcatcgctcttctcattagcgaactgacacgc 300
DB 245 ATCTTAATCAGTACTCTGCTTACATCTGCATCGCTTTCTCATTAGCGGAAGTACACGAC 304
QY 301 gctacccgactgcgggtggcaaatatcttttgcctcgtatcctcctggcaacaaatcaatc 360
DB 305 GTCTACCGGACTGCGCGTGGCCAAATATCAATTTTGGCTCGATCTCTGCAACCAAAATCAATC 364
QY 361 aatcggagacatttatacagtgctggaactcgtgtctgtcttcttcatggatcgctatcgga 420
DB 365 AATCGGAGCAVTTCAATACGTTGCGGACTGCTGCTGCTTTTCATGATCGCTATCGGA 424
QY 421 agctcagtgacatgatacctgtcgaagatcccgccgctcgtatagcgcctatagtaac 480
DB 425 AGCTCAGTGACATGATACCTGTCTCAACAGATCCCGGCGCTGATAGCCGCTTATAGTCAAC 484
QY 481 acatactcccaggattcgtggcatgtcttctcctcatctacagaggagtcgctgtgtgtg 540
DB 485 ACATACTCCAGGATTCGTGGCATGTCTTCTCATCTACGAGGGAGTCCGCTGGTGGTG 544
QY 541 ctcttgctcaactgttctgcccgaagaaaccccttggttcatgaatcgattcggc 600
DB 545 CTCCTGTCAACTGTTTGGCCCTGAAAGAAACCCCTTGGGTTCAATGAATCGGATTCGGC 604
QY 601 ctcaagatcgctctcttgatctccttcttctcgcacatctcagcggcgtccaaaccccaag 660
DB 605 CTCAGATCGCTCTTTCGTGATCTCCTTTATCGCCATCTTACGCGGTTCACACCCCAAG 664
QY 661 gctccaaactcacaggatgactgcttgagcaactactgctggtggtccgacggctc 720
DB 665 GCTCAAACTCACAGGTATGGACTGCTTGGAGCAACTATATGCTGCTGGTCCGACGCGTC 724
QY 721 tgcctcatctggcccttgacatcctgcttctcatgttctgcttgagcagcagaatg 780
DB 725 TGCTTCATCCTGGGCCCTTTCGACATCTCTGCTTCTCATGTTCNTTGGCTTGGACGACGCAATG 784
QY 781 catctggctgaagaatcacagatgctgctgctgctacggttaccocaaagcagtgctgctgca 840
DB 785 CATCTGGCTGAAGAATCACAGATGCTGCTGCTACGGTACCCAAAGACGTGCTCAAGTGA 844
QY 841 atcataatgctctcgcacgccttcccatatacaaatcgcaattctgtatggaattaca 900
DB 845 ATCATAAATGGCTTCTGACCGCCTTCCATATACAAATCGAGTTCTGTATGGAATTACA 904
QY 901 gatctcgactctattctaagtctcgcgcggctatattccattcagagacaatgacgagctcc 960
DB 905 GATCTCGACTCTATCTTAAGTTCGCGCGGCTATATTCATTCGAGACAATGAGGCAAGTCT 964
QY 961 ctctcggtcgtcagtttgtaacgggtcctctcctcgttggcggtgctgtagtgccctcttc 1020
DB 965 GCTCGAAATTCGGCTTTTGTGAACGGTCTCTCATGTGGCGGTATGCTGATGGCTCTCTTC 1024
QY 1021 gccctcaacgctgtacaagagactgctctcgactcaactcagcttcttccgggagcaat 1080
DB 1025 GCGCTCAACGCTGTACAGAGACTGCGCTCTCGACTCACCTGGAGCTTCCCGCGGACAAAT 1084
QY 1081 gggctggtattttccactcatctcgaaacgcgattcattcccccgcgtggcgaagtctctgttg 1140

DB 1085 GGGCTGGTATTTTTCACATCATCTCGAACGCATTATCCCCCTGGCAAGTTCTGTTTGG 1144
QY 1141 tctctattcgcgacctggggaattcttgccacatcgcgatgtatatatttcttaggttctagc 1200
DB 1145 TCTCTATTCCGACCTCGGGAATTTCTGGCCACATCGCATGTATATTCTTAGGTCTTAGC 1204
QY 1201 aaagctttcaatgcttggtcaattcccgcttcttactccagcaactctctctctgac 1260
DB 1205 ACAGCTTTCAATGCCCTTGGTCAATTTCCCGCTTGTACTCCAGCAACTCTCTCTCTGATC 1264
QY 1261 ccaatcgccctactcctctaccocaaagcgagatccaaagtctcttgcgagacactcgtgct 1320
DB 1265 CCAATCGCCCTACTCTCTTACCAAAGGAGATCCAAAGTTCTTGGCCGAGCACTCGTGT 1324
QY 1321 ttgtgttaccgctgggaatcggttctctggtcaatgctgtagcgggtggttcttaccgctc 1380
DB 1325 TTTGTGTTACCGCGTGGGAATCGGGTTTCTGTGTCAATGTGCTAGCGGTGGTCTTCCAGCTCC 1384
QY 1381 gtcacacatgctgttttcaagcttcccaactgacctgacctacgcgcgcgtcaacacataat 1440
DB 1385 GTCCACACATGTGTTTTTTCAGCTTCCCACTGACCGTGCCTACGGCGGTCAACCATGAAT 1444
QY 1441 tacacaagtcgattatagcgcttgcaactgctctgtgtcttgaactgggtgctgcat 1500
DB 1445 TACACAAGTCCGATATATAGCGCTTGCACCTTGTCTTGGTGTCTTGAACCTGGGTGCTGCAT 1504
QY 1501 gccaggaagcattatcagggaccctcacttgagcttgagcgggtcgctcgagagcagaa 1560
DB 1505 GCACGAAGCATATATCAGGACCCCACTTGGAGCTTGACGGACGGTCTCGGAGACGAA 1564
QY 1561 ttccaagttggcccatga 1578
DB 1565 TTTCAAGTTGGGCCATGA 1582

RESULT 3
SCYGL077C/c 2475 bp DNA linear PLN 11-AUG-1997
LOCUS S.cerevisiae chromosome VII reading frame ORF YGL077c.
DEFINITION 272599 Y13135
ACCESSION 272599.1 GI:1322592
VERSION 272599.1
KEYWORDS baker's yeast.
SOURCE Saccharomyces cerevisiae
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 2475)
AUTHORS Rieger M., Mueller-Auer, S., Brueckner, M. and Schaefer, M.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2475)
AUTHORS MIPS.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-1996) Data collected by MIPS on behalf of the
European yeast chromosome VII sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
Martinsried, FRG; E-mail: Mewes@mips.embnat.org
FEATURES
Location/Qualifiers
1..2475
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="VII"
gene complement(366..2057)
CDS complement(366..2057)
/gene="HNMI"
/note="ORF YGL077c"
/codon_start=1
/protein_id="CAA96782.1"
/db_xref="GI:1322593"
/db_xref="SWISS-PROT:P19807"
/translation="MSIRNDNAGSGTWPQSSNASMHKRDRLRVEEIKPLDDMSKG
AVAADGEVHLRKSFLWSILGFGFLTNSWPGI1STSWAGI1SSGGPMI1VIGI1VAL

ISICIGTSLGESSAY PHAGGOFWMSLK LAPKYKREAAVMCGSFAYAGSVFTSASTT
LSVATEVVMYALTHPEFIPKRWHIFVCFELLHLFAMFNCYKGLPII SSSSLYSIL
LSFFTITLVLAACHGKFNDAKFVPATFNNEFKNGGIAFIVGLINPAWSSCLDCA
TMAAEVKEPPEVVIDIA LMGVVAIGEVTSYVIAEFSTODLDAYLSSTGAPILDI
YQALNGNSGAIFIGCLLIFTSFGCVIACHTWQARLCWSPARONGLPULRWSQVNP
TYPLINAHLMSCAWITLGLLYLASSTOSLITGAFLLLSIIPVICLLAKRNI
AHGFWLKGFGFSNIVILLGWTFVSFFPVPVPTKDNMNVYCVIVGYTAYSIL
YWKYKKEFHAESENEQAEYSNNFDITDSRFSVAASDVELENEHVPWKK"

BASE COUNT 871 a 481 c 444 g 679 t

Query Match 4.2%; Score 66.4; DB 8; Length 2475;
Best Local Similarity 43.3%; Pred. No. 5.5e-08;
Matches 470; Conservative 0; Mismatches 606; Indels 10; Gaps 3;
Qy 212 cggggggccgcgcacatctctcttctacggcatctcaatcagctactctcgtctacatctgca 271
Db 1798 CTGCTGGACCCATGATGTTGTTACGGTATAATTAATTTGTCCTGATTTCTATTGCA 1739
Qy 272 tcgctttctcattagccgaactgaccagcgtctaccgactgcggtggccaatatcatt 331
Db 1738 TTGCTACCTGTTGGGTGAATATCTTCGCATATCCGACGCCGGTGGTTCAGTTTGGT 1679
Qy 332 ttgcgtcgatccctggcaccacaataatcaatcggagcatttcatacagtgctgcggactcg 391
Db 1678 GGTCTTTGAAGCTTGGCCCTCCTAAAGTACAAAAGATTTTCGGCCCTACATGTGCGGTTTCAT 1619
Qy 392 tgcgttgcctttcattcgctcattcggaagctcagtgaccatgatacctgctcaacaga 451
Db 1618 TTGCTTACGAGGATCCGTGTTACAAAGCTTCAACTACGTTATCCGTTCTACCCAG 1559
Qy 452 tccggcgctgtagcgcctcattagtcacacatacctccagattcgatgcattctctcc 511
Db 1558 TGGTGTGATGTACGCTCTGACGCCCTGAAATTCATCCAAAGATGGCATATATCG 1499
Qy 512 tcatctacgaggagtcgcgtcgtggtgctctcttcttcaact----tggttgcctgaa 566
Db 1498 TTTGCTTTGAATTTGTTGCAATTTGTTCTTGTATGTTTTTCAACTGTTTACGGTAAATCCTTAC 1439
Qy 567 aagaaaccttgggttcataagattcggatcgcctcagatcgctctctctcgtgactc 626
Db 1438 CTATCAATTTCCTCTCTATACATTTCCCTACTATCTCTTTTTCACAAATTA 1379
Qy 627 ctttatcgcatcttagcgggtccaaaccccaagctcacaactcaaggtatgactgc 686
Db 1378 CTGATTGGCATGTTCTCATGA--AAGTCAACGATGCAAGTGTGTTTTGCGCACAT 1321
Qy 687 ttggagcaactactggtggttcgcagcgggtctgctctcctcctgggctcttcgacatc 746
Db 1320 TAATAATGAACAGGTTGGAAGACGGGGTATCGCCCTTTATGTCGGTTTGATTAACCC 1261
Qy 747 ctgcttcatttcttcgttgagcgcagcaatgcattcgtcgtgaagaatgcacagatgc 806
Db 1260 AGCTTGGTCATTTTCGGCTTGACTGTGCAACCCATATGCGTTTGAAGTTGAAAACC 1201
Qy 807 tgcctgacggtaccacaagcagtggtcagtgcaatcaatattggtcttcgcacgcctt 866
Db 1200 AGAAGAGTATTTCCCATCTGATCATGGGAACAGTCGCCATTTGGGTTGTGCACTTCCT 1141
Qy 867 tccatatacatcgcgagttctgtatggaattacagatcgcactctatctcctaagttccgc 926
Db 1140 TTGTTATGTTATCGCTATGTTCTCTCTATACAAAGATCTGACGCTGCTCTTCTCTAC 1081
Qy 927 cggctatatccca---ttcgaacaataacacagcctccctcgtcgtcagtttggcaac 983
Db 1080 AACAGGGCCCCAAATCTTGACATTTATATCAGGCATTTGGGTAAATAATCAGGTGGCAT 1021
Qy 984 ggtctctcattcgtgcggtatcgtggtcctctcttcgcccctcaacgctgtacaagagac 1043
Db 1020 TTTCCCTGGGTTGCTGATTTCTATACCTCTTTTGGTTGCGTGATTCGTTGTCACACTG 961
Qy 1044 tgcgtctcagctcaccctgagcttttgcgggggacaagtggtggtgatttttccactcatct 1103

Db 960 GCAGGCAAGGTTATGTTGGTCATTTGCCAGGCAATGGTCTTCATATCCCGTTATG 901
Qy 1104 cgaacgccttcatcccccgcgtgcgaagtctcctggtctctctctctcgcacgtcgggaat 1163
Db 900 GTGCAAGTTAAACCCACACATGTTGTACCTTTGAAGCCTCATTTAATGTATGTCGCGCTG 841
Qy 1164 tctgcccacatcgcgatgatattcttaggtctctagcacagcgttcaatgccttggctca 1223
Db 840 GATAACCCCTCATTTGGCCTACTTTATTTGGCTTCCAGTACGCGCTTTTCAGTCTTAATAC 781
Qy 1224 ttcgcgccttctaccagcaactcctcctcctcctcctcctcctcctcctcctcctcct 1283
Db 780 AGTTGTATTCATTTTATTTGTTATCTACATCATTCCTCGGTATATGTTACTTCTCTAA 721
Qy 1284 aaagcg 1289
Db 720 AAAGCG 715
RESULT 4
LOCUS YSCCTR 2725 bp DNA linear PLN 27-APR-1993
DEFINITION Saccharomyces cerevisiae choline transport protein gene, complete cds.
ACCESSION J05603.1 GI:171329
VERSION choline transport protein.
KEYWORDS choline transport protein.
SOURCE S.cerevisiae DNA.
ORGANISM Saccharomyces cerevisiae
Eukaryote; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Nikawa,J.-i., Hosaka,K., Tsukagoshi,Y. and Yamashita,S.
1 (bases 1 to 2725)
Primary structure of the yeast choline transport gene and regulation of its expression
J. Biol. Chem. 265, 15996-16003 (1990)
90368823
Draft entry and computer-readable sequence for (J. Biol. Chem. (1990) In press) kindly provided by J.-I. Nikawa 30-JUL-1990.
FEATURES
Location/Qualifiers
1..2725
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
446..>2211
/gene="CTR"
/product="choline transport protein"
446..2392
/gene="CTR"
455..2146
/gene="CTR"
/codon_start=1
/product="choline transport protein"
/protein_id="AA34537.1"
/translation="MSIRNDNASGGYMQPDQSSNASHKRLRLVREEIKPLDDMSKG
AAADGEVHLRKSFLWSILGVGFLTNSWFGISTSHVAGISSGGPMHIVYGIIVAL
ISICIGTSLGESSAY PHAGGOFWMSLK LAPKYKREAAVMCGSFAYAGSVFTSASTT
LSVATEVVMYALTHPEFIPKRWHIFVCFELLHLFAMFNCYKGLPII SSSSLYSIL
LSFFTITLVLAACHGKFNDAKFVPATFNNEFKNGGIAFIVGLINPAWSSCLDCA
TMAAEVKEPPEVVIDIA LMGVVAIGEVTSYVIAEFSTODLDAYLSSTGAPILDI
YQALNGNSGAIFIGCLLIFTSFGCVIACHTWQARLCWSPARONGLPULRWSQVNP
TYPLINAHLMSCAWITLGLLYLASSTOSLITGAFLLLSIIPVICLLAKRNI
AHGFWLKGFGFSNIVILLGWTFVSFFPVPVPTKDNMNVYCVIVGYTAYSIL
YWKYKKEFHAESENEQAEYSNNFDITDSRFSVAASDVELENEHVPWKK"
polyA_signal
/gene="CTR"
2387..2392
/gene="CTR"
BASE COUNT 806 a 469 c 545 g 905 t
ORIGIN


```
Db 17114 KGMVCHYCWNGRGVCSCMTSRSAMCCSYAKCKSMCYCYGMSMKGYIYACSYRG 17173
QY 634 gccatttagcgggtccaaacccaagggctccaaactcacaggtat---gacgtgcttgg 690
Db 17174 SMSKSKYMRGTVSTGCGCCCTTTTCCCCCNANTGGGAAGCTTTTNCNKTKTYSYRKR 17233
QY 691 agcaactatactggtgctgcagcggcgtctgcttcttctatcctggccttccacacccgc 750
Db 17234 NGCAMPKYNVYNNMRSRSCACGMSCTKYKSSMTMSMASYWCWMSYCYSMRSMASRCMS 17293
QY 751 tteatgttccattgcttgacgcagcaatgcattgctgctggaagaatgcacagatgctct 810
Db 17294 WSSYKMKMSRWKSCYKCCWCKMCKYCMRSMRSGMSYMYASWKSRSRGCTCTRCY 17353
QY 811 cgtacgttacccaaagcagtggtcagtgcaatacataatggtcttctgacacgcctttcca 870
Db 17354 WCMSSKSCYKSYNNMRSRCKMKWRCWGSWGSASRSCKYKCYKMRCSMMSSKCYR 17413
QY 871 tatacaatcagcttctgtatgaattacagatctcagctctctctctctctctcgcgcgc 930
Db 17414 CAGCMWKKGYMYRWCWMSKRWKGSAM----YRMRMKWKGAMWCKCYSRMSR 17468
QY 931 tatattccattcagacgaatgcagcagtcctcctgctgctcagtttgcacaggtctctc 990
Db 17469 CMWKKSYACSCRCAMMSCTYMKCASYCMCSYWCCTCTCTCTCTCTCTCTCTCTCTCT 17528
QY 991 tcattgctgctgctgctgctgctctctctcctcctcctcctcctcctcctcctcctct 1050
Db 17529 KSKNSWSSMSYKGRKSKYSMCTSRGAMSCWRRCCYMRGASSMRPAGSMRRAKGRSW 17588
QY 1051 cgactacacctggagcttgcgcgggacaaatggcgtggtgatttccactcactcactcgaacgc 1110
Db 17589 GGRSKWMTGCMWRSKYIYCTGRRMMTYMCCWRRRRYMYRSMAMGMRKSSWGMW 17648
QY 1111 attcattccctgctgctgcaagctctctctgctctctctcctcctcctcctcctcctc 1170
Db 17649 MMSASRRCKSASRSWCSMRKMGWRCWGSWCKMWSRSASCKCKGSRGMRRSKR 17708
QY 1171 acatgcgagtgatatt 1187
Db 17709 SSKYRGRGKKRSNTK 17725

RESULT 8
SPAC1039 SPAC1039 27893 bp DNA linear PLN 15-DEC-1999
LOCUS S.pombe chromosome I cosmid c1039.
DEFINITION A1133521
ACCESSION A1133521
VERSION A1133521.1 GI:6594259
KEYWORDS 2,2-dialkylglycine decarboxylase; amino-acid permease;
Aminotransferases class-III; Bacterial transferase hexapeptide;
C3H2 type; esterase; family 31 glucosidase; Glycosyl hydrolases
family 31; lps; pyridoxal-phosphate; serine acetyltransferase;
sexual differentiation process putative amino-acid permease isp5;
sugar tr; translation initiation inhibitor; transporter; UPP0076;
zf-C3H2; zinc finger protein.
fission yeast.
SOURCE Schizosaccharomyces pombe
ORGANISM Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomycetes.
REFERENCE 1 (bases 1 to 27893)
AUTHORS Hunt,C., Aves,S., McDougall,R.C., Rajandream,M.A. and Barrell,B.G.
TITLE Direct Submission
JOURNAL Submitted (13-DEC-1999) European Schizosaccharomycetes genome
sequencing project, Sanger Centre, The Wellcome Trust Genome
Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
and Department of Biological Sciences, University of Exeter, Perry
Road, Exeter EX4 4QG, United Kingdom
Notes:
Details of yeast sequencing at the Sanger Centre are available on
the World Wide Web.
```

(URL, http://www.sanger.ac.uk/Projects/S_pombe/)
During 1995 to 1996 about 66% of S. pombe chromosome 1 was sequenced by the Sanger Centre. The sequencing of the S. pombe genome is now being continued with funding from The European Commission, fourteen European sequencing laboratories, including the Sanger Centre, are participating in the project.
The protein coding regions (CDS) have been predicted with the help of computer analysis using the Genefinder program in PomBase (an ACEDB database) with additional predictions for the branch-acceptor sites supplied by the program Splice. CAUTION: It is possible that for any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites.
CDS are numbered using the following system eg SPBC35H2.01c.SP (S. pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.
Cosmid c1039 is overlapped at the 5' end by cosmid c29B12, EMBL entry SPAC29B12, accession number 299164 and at the 3' end by cosmid c922.

FEATURES
source
1. 27893
/organism="Schizosaccharomyces pombe"
/strain="972h"
/db_xref="taxon:4896"
/map="IR"
/clone="cosmid c1039"
1. 49
/note="nominal overlap with cosmid SPAC29B12 S. pombe chromosome 1"
1375..3078
/gene="SPAC1039.01"
1375..3078
/gene="SPAC1039.01"
/note="SPAC1039.01, len:567,
SIMILARITY:Schizosaccharomycetes pombe, 074537, putative
amino-acid permease., (557 aa), fasta scores: opt: 2304,
E():0, (62.5% identity in 520 aa)"
/codon_start=1
/label="SPAC1039.01
/product="amino-acid permease"
/protein_id="CAB63537.1"
/db_xref="GI:6594260"
/translation="MSSIMERTVSSSKHSTFPSPSDEVCKAAEIVDTHENFYDDQG
DIDNPEELAEGLYKQEFRRLSLFGISFSLVGLMPLPSVASTLVGLAWYVYGLLW
AMLIAMFFLVCVSMARICAMPTSGLYAAVAFAPKGMPLASHTGWSYICNI
IGQPSVSAASMLIGAVTVNRPFVQRMOWFLAVAIQCFNCVLACLPTRISRN
GVATYNTAFLFAGITLAVGKNHNFVKGTIGVDYINTQWPTGFAILLFSNP
WMSGYDAPHLSEPCSNASNAKAVMTAVIGVVGWIMQIIIVATLNDIVSNT
SCSMNTAVLCAMPKAKGLILTLISAILMGOSALIASRIAYSVARDGILPFSGW
IGTVNPTQTPNVAVICNIIISILILFTAGTTLDAVFSVGAVAFIAPTVPPIAIR
VEFTKDFRRGPNWNLGRSPGLLAVSFVALMIPILCFPSVKNPQAQMNWTCILY
GGPMLETLTVAISARKWFKGPKASAHYKRPGEESDIVEGVQADIPSSSDQLKING
L"
1543..2964
/gene="SPAC1039.01"
/note="pfam match to entry PF00324 aa_permeases, Amino
acid permease"
5742..7547
/gene="SPAC1039.02"
5742..7547
/gene="SPAC1039.02"
/note="SPAC1039.02, len:601,
SIMILARITY:Schizosaccharomycetes pombe, 013781, hypothetical
71.4 kd protein cl796.03 in chromosome i precursor., (635
aa), fasta scores: opt: 936, E():0, (37.7% identity in 605


```
/gene="P0660F12.8"
32084. 34153
/gene="P0660F12.8"
note="probably inactive due to frameshift in CDS
pseudogene
similar to Oryza sativa chromosome 10, OSJNBA0046L02.1"
/pseudo
join(37010. 37042,37855. .38133,38337. .38756)
/gene="P0660F12.9"
join(37010. 37042,37855. .38133,38337. .38756)
/gene="P0660F12.9"
note="contains EST AU102097(C10257)
unknown protein"
/codon_start=1
/protein_id="BAB63845.1"
/db_xref="GI:15290155"
/translation="WSTSLPAERKAARKALKGWDPTIKRNFPPYWRSLGAQLSV
SNQADERSHLRPFQFRNSVYVSNESLSNHNISFLGRIOQLSGTHRESRYSTAFRML
TKHVAVDNHRSAKYIAARVFCGLKLVANIHVTKTLFRVMLSTLNDKQHVACCI
RLPSTREMTCPVACCTCFRNVSLNTEQVKVAKHFSLSAISSTYIRVMILECYW
ATWPKELAKHHQNECHVSIRYKLL"
join(42085. 42216,42290. .42497,43572. .43702)
/gene="P0660F12.10"
join(42085. 42216,42290. .42497,43572. .43702)
/gene="P0660F12.10"
note="contains ESTs C27575(C52254),AU082350(C52254)
unknown protein"
/codon_start=1
/protein_id="BAB63846.1"
/db_xref="GI:15290156"
/translation="WSTSEAGLPDPATAASSLVDCRLPGRCEPLHSSPLSHLH
VDVPLFGSTLLSTVBLTSPCRSPPLGWRGARAANAAPKPRKAWAKYGD
DGSKATMARIYQVNGEQRCKEIEDTKHLESGSITDPKAKLAFSDIDKDSR"
complement(45916. 46414)
/gene="P0660F12.11"
note="probably inactive due to frameshift in CDS
pseudogene"
/pseudo
complement(45916. 46414)
/gene="P0660F12.11"
complement(join(52037. 52146,52966. .52990,54211. .54418,
54685. 54725,55362. .55628,55945. 56234,56567. 56644,
57192. 57264,57957. 58041,59866. 60233))
/gene="P0660F12.12"
complement(join(52037. 52146,52966. .52990,54211. .54418,
54685. 54725,55362. .55628,55945. 56234,56567. 56644,
57192. 57264,57957. 58041,59866. 60233))
/gene="P0660F12.12"
note="hypothetical protein"
/codon_start=1
/protein_id="BAB63847.1"
/db_xref="GI:15290157"
/translation="MDPRLLSAARRGDREWLAVLGGAAAPADQGAVDVAGPPDPGQ
APAPPPPSAAALLLDVATTPRCDLSALHVAVAGSDDFLCARMYIHAEGILGH
TNATGDADEEARLRKRFRFSQPTAREIKIFTCSAVITITALLGPVSTDSMTV
RGISGLAVKCYSDNLREKREPERPSGAALQWLQGVGDGDEGRWSMVCRGEEV
GLGGDVRKATAPAVACDWSATAAHTRGGRGVGRWSARHSDGLPWLARQRRQA
YARHAPRRRQPKQKGTPTKDLRLAICRLTGNRGCRISSHRNTKTYHPLPTIRP
EPGLRLRNHHSYTEYGRIPRYGHSGSSPDSSTIPARAGEASTGRAPYLEDDSR
LRRRLARLACWGSSEKRVQLGAEAGDAVDTRWPNQIONTYAHIASRGDVGARVERD
```

Query Match 2.6%; Score 41.2; DB 8; Length 147203;
Best Local Similarity 49.5%; Pred. No. 2; Mismatches 108; Indels 0; Gaps 0;
Matches 106; Conservative 0; GCATCATCTCCGTGATCGCGCGGTGACGACGACGT 119499

QY 134 tcaccatgtctcttggccttggatggaacagtggtgggtggtatctctcaggcagtc 193
DB 119558 TCTCCAACTTCGCTTCTCTTCATCATCTCCGTGATCGCGCGGTGACGACGACGT 119499

QY 194 tcacgtcgcctactacggggggggccgctcactctcttaccgcatctcctaactagta 253
DB 119498 ACAGCAGCGGGCTCCGCTACGCGGGCGCGGTGTCATCAGCTCGGCTGCTGCTGT 119439

QY 254 ctctggtctacatctgccttctctcattagcgaactgacccagcgtctaccgactg 313
DB 119438 CCGCCTTCAACGGGTGCATGGCCCTGTCATGCGCAGATCTGCTCCGCTACCGGACCT 119379

QY 314 cgggtggccaatatactatttgcgtcgatctctggc 347
DB 119378 CCGGTGGCTCTACTACTACTGGAGTGCCCAAGCTGCG 119345

RESULT 10
AF429315/c
LOCUS AF429315 125020 bp DNA linear PRI 18-JAN-2002
DEFINITION Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION AF429315
VERSION AF429315.1 GI:17646244
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 125020)
AUTHORS Holmes, S.E., O'Hearn, E., Rosenblatt, A., Callahan, C., Hwang, H.S.,
Ingersoll-Ashworth, R.G., Fleisher, A., Stevanin, G., Brice, A.,
Potter, N.T., Ross, C.A. and Margolis, R.L.
A repeat expansion in the gene encoding junctophilin-3 is
associated with Huntington disease-like 2
Nat. Genet. 29 (4), 377-378 (2001)
JOURNAL MEDLINE 21583737
PUBMED 11694876
REFERENCE 2 (bases 1 to 125020)
AUTHORS Holmes, S.E., Ingersoll-Ashworth, R.G., Ross, C.A. and Margolis, R.L.
Direct Submission
TITLE Submitted (05-OCT-2001) psychiatry, Johns Hopkins Medical
JOURNAL Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES
source
1. 125020
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q24.3; between D16S520 and WI-12410"
/note="Isolated from a patient with Huntington's
Disease-like 2 (HDL2).
complement(35581. 35746)
repeat_region rpt_type=tandem
rpt_unit=ctg
complement(36507. 36887)
/gene="JPH3"
/product="Junctophilin 3"
/gene="JPH3"
/product="Junctophilin 3"
/codon_start=1
/protein_id="AAL40941.1"
/db_xref="GI:17646245"
/translation="MSSGRFNFDDGSVCGGWEDGKAHGHGVTGPKGGVYTGWSW
HGPEVLGVYTWQSGTYCTQACGRHIGLESKGVYKGVTHGTFKGRYGVRCAG
NGAKYEGTWSNLQDGYGTETYSQ"
BASE COUNT 29056 a 32731 c 30696 g 26283 t 4254 others
ORIGIN

Query Match 2.6%; Score 41; DB 9; Length 125020;
Best Local Similarity 11.8%; Pred. No. 2.2;
Matches 96; Conservative 350; Mismatches 361; Indels 6; Gaps 3;
QY 32 aaggcgggacagcagcaacagacagacagcagcgcgaggtggtgctccg 91
DB 51567 RASRCKGNGKMAKSCMRMYMYKGGGAGKCCSAGGTMGTGTSRWSACCMKMSKASKA 51508

```
QY 92 agtccctgaacgtctctctgagagaagaacaatttggcacacatcacaccatgctgcttgg 151
: : : ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 51507 RYATAGTAAGSMCGTCYKMKCTRRMATAACAMHRWAMWMSCCMRKCTWKSXMGRR 51448
: : : ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 152 ccttggattgaacagctgggtggtatctcaggcagctctccagctgcgcctactag 211
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 51447 MRYCWSARWYVAGSYMKCRSGYKGRSRSGMSWTRGYSYSAKYMSKSGAGA 51388
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 212 cggggggcccgctcactctctcttaccgcatcctaatcagctactctcgtctacatcga 271
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 51387 KCKGAGSYAKGSNGRWGSYKRWAWYSCWMSASKSMTCYWSCCTCKGCVWYWSW 51328
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 272 tggcttctcattagcgaactgacagcgtctaccgactgcgggtggccatacatt 331
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 51327 KYGAGCMWYHWCCTCMARWARRARMAARRAMWMAAMWMMWMMMBHAAAMW 51268
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 332 ttgctgcatcctggcaccacaaatcaatcaatcggagcattctcactcgtgctgagactcg 391
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 51267 ARWAAAKCWYYSYKYCWMSMMWASWAAAMWMMCTAAWYKKW--TRMRWWSYSSS 51210
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 392 tgcgttgccttctcagctgctcactcgggaagctcagtgacatgatacctgctcaacaga 451
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 51209 YAYAWYATMCWYCYSSMYAARMGTYSGGKKHKAWSCTCCRRGCCMKKGMWCS-- 51150
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 452 tcccgcgctgtagcgcgctctatgacacatactcaccaggtcgtgcatgtcttcc 511
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 51151 TTYTGMSCGKSRWMTSMACWYWGSRRTKSMATSCMRGAMRSGGMRKYWSCRKY 51092
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 512 tcatcagaggagtagtcgctggtggtctcttcttgaacttcttggcctgaaagaa 571
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 51091 GMRGRWSYKGCSSYMSMGYSGRTSYKYSYTGCGYKCSAKMCKSKSTKSCCTKX 51032
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 572 acccttgggtctcgaatcggatctggcctcagctcgtctctcttctgtaactccttta 631
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 51031 YSCSMGRYSYCTCWMGRMCKMK--SSCCASRSCKYKMGWSMCRGRGMSCWYSS 50974
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 632 tgcactctagcagctcacaacccaagctcacaactcagctgtagactgactga 691
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 50973 MKSYMSASYSSAGRMCKYKTSCKMYYSMBHKSDBBGYSYKSYBYWTCMKKDS 50914
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 692 gaactactgctggcgcagcgcgtctgcttctcactcgttggccttgcacatcctgct 751
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 50913 MCWSDSHMDMMYDKRMKVSSTGRYHBSAWMKSMRSDSDSGYSGMYHMKSYKS 50854
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 752 tcatcttctggtggcagcgaatgcatctgctggtgaagaatgcacagatgctgctc 811
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 50853 SASKHMYSGMKSSVYBMDCMSYTSKSKSVRMBRRWSGSYBRYRKCVMRYBHRSTR 50794
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 812 gtacggtaccacaagcagtgctcagtgcaatca 844
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 50793 SYHVGKRSQWVRDSVSRSSVYTSWSSWRCR 50761
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 11
AF053231 Neurospora crassa 1996 bp mRNA linear PLN 28-MAR-1998
LOCUS Neurospora crassa amino acid permease (aap2) mRNA, complete cds.
DEFINITION AF053231
ACCESSION AF053231
VERSION AF053231.1 GI:2995850
KEYWORDS
SOURCE Neurospora crassa.
ORGANISM Neurospora crassa.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.
1 (bases 1 to 1996)
Margolles-Clark,E. and Bowman,B.J.
Direct Submission
TITLE Neurospora crassa amino acid permease (aap2) mRNA, complete cds.
JOURNAL Cruz, 444 Sinsheimer Laboratories, Santa Cruz, CA 95064, USA
Location/Qualifiers
FEATURES
1..1996
source /organism="Neurospora crassa"
```

```
/db_xref="taxon:5141"
/chromosome="LGI"
/map="close to R15.3, AP3la.3"
1..1996
/gene="aap2"
/gene="aap2"
100..1755
/function="amino acid transporter"
/notes="similar to GABA permeases; specificity unknown"
/codon_start=1
/product="amino acid permease"
/protein_id="AAC08355.1"
/db_xref="GI:2995851"
/translating="MSFSPNKSADATQITETROGTPSSGEAAASTSTSTESGD
KALEALGYTPVFKBESRMSFSFAVSIQVYGTLMSTWYGLQAGGAALIMSWIIG
GAGWALAYSIAEIASAYSSGAMYFKLAPREOYPLCHTAGYLVNLCVAGGAS
TEYASQMLAAVSIITSNESYVPTPHVVGVIHGLTIHAMIWTLTAHLNRLTSGYV
VFHISVLGGACVTLVOKRHHDLKYATFNQPSGWSPPGFALFGLTFLAMITG
CDGPARIAEAKNPQMVVPRAIANATTFTYVIGFTFNLVVCMDKDLNPSGQP
VAQLFNGMRAPAIFFTLGFGVMNLVAIPGMOAGSKTIFALSRDLDFSHIWRV
SKRSQTPLIATWYIAVLEIIINLLGLASTAIGAVNCTVALNVSYPFLICKWYIG
RMQSGPMHMGKYSWVYNFAVANMTFMVIFFTFVRVTPPENMNTAIVVFFFLILA
LVFWYTHGRHYVTGPLTHSPRATMSVTRPMSVIGTCTTL"
BASE COUNT 423 a 539 c 477 g 557 t
ORIGIN
Query Match 2.58; Score 40; DB 8; Length 1996;
Best Local Similarity 51.7%; Pred. No. 3;
Matches 91; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
QY 1350 ggtcaatgctagcgggtgcttcacgtccgtccaccatgctgttttccagcttcccaact 1409
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1494 GGTCAATGCGTTGGCGGTTCGCTGGAATAGTTCATGGCGGTGATCTCTTCTCCCCAC 1553
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1410 gaccgtgcctacgcgcgcgtcaaccatgaattacacaaagtcgcattataggcgttgcaact 1469
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1554 GCGSGTACCTGTTACACCTGAGATATGAACACGACCATAGTGTCTTTCTCGTCTCT 1613
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1470 tgccttgctgctgaactggcgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1525
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1614 CATTTAGCCCTGCTTCTGTCACACCCATGGACGCCACTACTATACCGGCGCCC 1669
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 12
CNS01D5D 720 bp mRNA linear PLN 03-SEP-1999
LOCUS Botrytis cinerea strain T4 cDNA library under conditions of
DEFINITION nitrogen deprivation.
ACCESSION AL116393
VERSION AL116393.1 GI:5831609
KEYWORDS cDNA library; nitrogen deprivation.
SOURCE Botryotinia fuckeliana.
ORGANISM Botryotinia fuckeliana.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;
Helotiales; Sclerotiniaceae; Botryotinia.
REFERENCE 1 (bases 1 to 720)
AUTHORS Bitton,F., Levis,C., Fortini,D., Pradler,J.M. and Brygoo,Y.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr,
78026 Versailles, France
REFERENCE 2 (bases 1 to 720)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (01-Sep-1999) Genoscope - Centre National de Sequencage :
CP 5706 91057 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT The cDNA library to be analyzed within the framework of this
project was created using a Botrytis cinerea strain which was grown
under conditions of nitrogen deprivation, which is the normal
situation for B. cinerea during its development on its host plant.
The library was produced in an oriented direction, in the pBSII
vector.
```

| | | | | | | |
|---|----------------------------|---------|--|--|--|--|
| FEATURES | Location/Qualifiers | | neuropeptide binding" | | | |
| | source | 1..720 | /codon_start=1 | | | |
| BASE COUNT | 144 a 173 c 160 g 242 t | | /product="head-activator binding protein precursor" | | | |
| | 1 others | | /protein_id="AAC63362.1" | | | |
| ORIGIN | 144 a 173 c 160 g 242 t | | /db_xref="GI:3719423" | | | |
| | 1 others | | /translation="MKMAYVRIHKVSHNNSLPELLFIAPVQFVTNVNASSLSVNP | | | |
| Query Match | | | | | | |
| Best Local Similarity 53.7; Score 39.2; DB 8; Length 720; | | | | | | |
| Matches 80; Conservative 0; Mismatches 69; Indels 0; Gaps 0; | | | | | | |
| QY 1311 cactcgtgcttctgttacccgctggaacccgggtttctggtcaatgctagcgggtggt 1370 | | | | | | |
| DB 514 CANGCATGCTTCATTTTCGNTGGCAGATTGGATTGCTATCAACGCTGTTACCGTGTG 573 | | | | | | |
| QY 1371 ctccacgctccgaccacactggttttccagcttcccaactgacgcgtgcctacggcgctc 1430 | | | | | | |
| DB 574 TTGGATTACTCTGGCCATCATCTTTCTGCGATGCCACCGCTATTCCGGTCACCTCTTC 633 | | | | | | |
| QY 1431 aaccatgaattacacagtcgactatag 1459 | | | | | | |
| DB 634 AACTATGAATACCGGAGTGTAGTTTGTG 662 | | | | | | |
| RESULT 13 | | | | | | |
| AF092920/c | | | | | | |
| LOCUS | | | | | | |
| DEFINITION | | | | | | |
| ACCESSION | | | | | | |
| VERSION | | | | | | |
| KEYWORDS | | | | | | |
| SOURCE | | | | | | |
| ORGANISM | | | | | | |
| REFERENCE | | | | | | |
| AUTHORS | | | | | | |
| TITLE | | | | | | |
| JOURNAL | | | | | | |
| AUTHORS | | | | | | |
| TITLE | | | | | | |
| JOURNAL | | | | | | |
| FEATURES | Location/Qualifiers | | gene | | | |
| | source | 1..5359 | /gene="HAB" | | | |
| misc_feature | 1..5359 | | /note="encodes head-activator binding protein precursor" | | | |
| | 1..5359 | | /note="HAB" | | | |
| sig_peptide | 253..4983 | | /gene="HAB" | | | |
| | 2219..3000 | | /product="head-activator binding protein" | | | |
| misc_feature | 3001..3123 | | /rpt_family="LDLR class B" | | | |
| | 3172..4101 | | /note="encodes EGF-like domain" | | | |
| misc_feature | 4102..4749 | | /rpt_type="dispersed | | | |
| | 4750..4818 | | /note="encodes fibronectin type III domain" | | | |
| misc_feature | 4819..4983 | | /gene="HAB" | | | |
| | 4987..5359 | | /note="encodes intracellular domain" | | | |
| misc_feature | 5317..5321 | | /gene="HAB" | | | |
| | 5317..5321 | | /note="HAB" | | | |
| CDS | 1684 a 900 c 1045 g 1730 t | | /note="signal for fast mRNA degradation" | | | |
| | ORIGIN | | | | | |
| Query Match | | | | | | |
| Best Local Similarity 56.1; Score 39.2; DB 3; Length 5359; | | | | | | |
| Matches 74; Conservative 0; Mismatches 58; Indels 0; Gaps 0; | | | | | | |

Db 5700 YSYDASTCYTSRGTCYTTBYSTBTBNCRCNCYSDSTYRANCYSSTYDATBNSTY 5646

Search completed: April 26, 2002, 23:09:34
Job time: 13563 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 02:58:44 ; Search time 4999.95 Seconds
(without alignments)
4259.682 Million cell updates/sec

Title: US-09-882-694A-7
Perfect score: 1578
Sequence: 1 atgactccagaccagtg.....aattcaagttgggccatga 1578

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estha:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 22 | 1.4 | 728 | 10 | BE272788 601104991 |
| 2 | 21 | 1.3 | 832 | 10 | BE299324 601119739 |
| 3 | 20 | 1.3 | 243 | 9 | AW408618 UI-HF-BM0 |
| 4 | 20 | 1.3 | 272 | 10 | BF933836 CM2-NT024 |
| 5 | 20 | 1.3 | 277 | 10 | BE244343 TCBAP2E03 |
| 6 | 20 | 1.3 | 285 | 10 | BM152000 TCBAP1E11 |
| 7 | 20 | 1.3 | 343 | 9 | AA644607 af74b07.r |
| 8 | 20 | 1.3 | 345 | 9 | AA629140 af60b06.s |
| 9 | 20 | 1.3 | 363 | 10 | BT316879 |
| 10 | 20 | 1.3 | 365 | 9 | AW407577 |
| 11 | 20 | 1.3 | 365 | 10 | BM152194 |
| 12 | 20 | 1.3 | 391 | 9 | AW222767 |
| 13 | 20 | 1.3 | 434 | 9 | AL120231 |
| 14 | 20 | 1.3 | 455 | 10 | BM174982 |
| 15 | 20 | 1.3 | 455 | 10 | BE241865 |
| 16 | 20 | 1.3 | 466 | 10 | RG395808 |
| 17 | 20 | 1.3 | 469 | 10 | BM153121 |

| | | | | | |
|----|----|-----|-----|----|----------|
| 18 | 20 | 1.3 | 471 | 10 | BF043417 |
| 19 | 20 | 1.3 | 482 | 10 | BM049825 |
| 20 | 20 | 1.3 | 490 | 10 | BT760416 |
| 21 | 20 | 1.3 | 538 | 9 | AW077262 |
| 22 | 20 | 1.3 | 583 | 10 | BG478113 |
| 23 | 20 | 1.3 | 586 | 10 | BE900139 |
| 24 | 20 | 1.3 | 591 | 10 | BM490083 |
| 25 | 20 | 1.3 | 606 | 10 | BE384337 |
| 26 | 20 | 1.3 | 620 | 10 | BE384039 |
| 27 | 20 | 1.3 | 630 | 10 | BG761867 |
| 28 | 20 | 1.3 | 634 | 10 | BE900587 |
| 29 | 20 | 1.3 | 644 | 10 | BG720272 |
| 30 | 20 | 1.3 | 648 | 10 | BT772780 |
| 31 | 20 | 1.3 | 656 | 10 | BT391871 |
| 32 | 20 | 1.3 | 667 | 10 | BT717889 |
| 33 | 20 | 1.3 | 667 | 10 | BF204617 |
| 34 | 20 | 1.3 | 675 | 10 | BF025655 |
| 35 | 20 | 1.3 | 679 | 10 | BE887718 |
| 36 | 20 | 1.3 | 682 | 10 | BT765819 |
| 37 | 20 | 1.3 | 683 | 10 | BT724330 |
| 38 | 20 | 1.3 | 688 | 10 | BE294290 |
| 39 | 20 | 1.3 | 690 | 10 | BT753384 |
| 40 | 20 | 1.3 | 724 | 10 | BE887301 |
| 41 | 20 | 1.3 | 729 | 10 | BT257589 |
| 42 | 20 | 1.3 | 733 | 10 | BE281295 |
| 43 | 20 | 1.3 | 740 | 10 | BF204036 |
| 44 | 20 | 1.3 | 744 | 10 | BT459520 |
| 45 | 20 | 1.3 | 752 | 10 | BG819515 |

ALIGNMENTS

BE272788 728 bp mRNA linear EST 13-JUL-2000
601104991F1 NTH_MGC_14 Homo sapiens cDNA clone IMAGE:3347735 5',
mRNA sequence.
BE272788
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 728)
NTH-MGC http://mgc.nci.nih.gov/
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE
Unpublished (1999)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov
Plate: LHCMI39 row: f column: 24
High quality sequence stop: 646.
Location/Qualifiers
source
1. 728
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3347735"
/clone_lib="NIH_MGC_14"
/tissue_type="renal cell adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: kidney; Vector: pORF7; Site_1: xhoI; Site_2:
EcoRI; cDNA made by oligo-dr priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of

```

California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
BASE COUNT      126 a      229 c      248 g      125 t
ORIGIN

Query Match      1.4%; Score 22; DB 10; Length 728;
Best Local Similarity 100.0%; Pred. No. 3 6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 503 atgtctctctcatctacgagg 524
|||||
Db 164 ATGCTCTTCCATCTACGAGG 185

RESULT 2
BE299324      832 bp      mRNA      linear      EST 20-JUL-2000
LOCUS
DEFINITION
601119739F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:3029577 5',
mRNA sequence.
ACCESSION
BE299324
VERSION
BE299324.1 GI:9183072
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 832)
AUTHORS
NIH-MGC http://mgc.ncl.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM89 row: n column: 10
High quality sequence start: 11
High quality sequence stop: 708.
FEATURES
Location/Qualifiers
1..832
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3029577"
/clone_lib="NIH_MGC_17"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```

BASE COUNT      148 a      232 c      276 g      176 t
ORIGIN

Query Match      1.3%; Score 21; DB 10; Length 832;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 503 atgtctctctcatctacgagg 523
|||||
Db 198 ATGCTCTTCCATCTACGAGG 218

RESULT 3
AW408618
LOCUS
DEFINITION
UT-HF-BM0-adv-d-06-0-UI.r1 NIH_MGC_38 Homo sapiens cDNA clone
IMAGE:3062843 5', mRNA sequence.
ACCESSION
AW408618
VERSION
AW408618.1 GI:6927675
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 243)
AUTHORS
NIH-MGC http://mgc.ncl.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 forward.
FEATURES
Location/Qualifiers
1..243
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3062843"
/clone_lib="NIH_MGC_38"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/lab_host="DH10B (LTI)"
/notes="Vector: p773-Pac; Site_1: NotI; Site_2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(2.5-3.5kb). Directionally cloned. Cells provided by Louis
M. Staudt, Ph.D. Library preparation by Maria de Fatima
Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT      35 a      69 c      102 g      37 t
ORIGIN

Query Match      1.3%; Score 20; DB 9; Length 243;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 505 gtctctctcatctacgagg 524
|||||
Db 200 GTCTTCTCATCTACGAGG 219

RESULT 4
BF933836
LOCUS
DEFINITION
CM2-NT0244-211200-642-e02 NT0244 Homo sapiens cDNA, mRNA sequence.
ACCESSION
BF933836
VERSION
BF933836.1 GI:12351160
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 272)
AUTHORS
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,I.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
```



```

/cell_type="pre-B cell"
/dev_stage="pediatric 2 years"
/lab_host="DH10B"
/notes="Vector: lambda pSB; Site_1: BamHI; Site_2: EcoRI;
First strand cDNA was primed with an anchored
XhoI-oligo(dT) primer [5'GGAGACTCGAGCGCGCGAGGAG(T)YN
3'; V-A,C,G; N-A,C,G,T] and then dg tailed. Second strand
was primed with a BamHI-dC primer
[5'AGAGACTCGATCGCGCGCGCAATAATAAT(C) 3'].
Double-stranded cDNA was then digested with BamHI and XhoI
and directionally cloned into the BamHI and XhoI sites of
lambda pSB vector. Library went through one round of
normalization. Library was constructed by Wei Yu at RIKEN
of Japan (Carninci P, Westover A, Nishiyama Y, Osumi T,
Itoh M, Nagaoaka S, Sasakini, Okazaki Y, Muramatsu M,
Schneider C, Hayashizaki Y, High efficiency selection of
full-length cDNA by improved biotinylated cap trapper.,
DNA Res 4: 1, 61-6, Feb 28, 1997)."
47 a 73 c 112 g 53 t
BASE COUNT
ORIGIN

Query Match 1.3%; Score 20; DB 10; Length 285;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 505 gtcttcctcatctacgaggg 524
|||||
Db 167 GTCTTCTCATCTACGAGG 186

RESULT 7
LOCUS AA644607
DEFINITION af74b07.r1 Soares_NHMPu_S1 Homo sapiens cDNA clone IMAGE:1047733
(HUMAN); mRNA sequence.
ACCESSION AA644607
VERSION AA644607.1 GI:2569825
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 343)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisler,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Lennon,G., Marra,M., Martin
J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wyllie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 828 Std Error: 0.00
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 336.
Location/Qualifiers
1..343
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1047733"
/clone_lib="Soares_NHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/notes="Organ: mixed (see below); Vector: pT7n3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NBHPU, and fetal heart NBH19W) were mixed, and as circles
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

Query Match 1.3%; Score 20; DB 9; Length 343;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 505 gtcttcctcatctacgaggg 524
|||||
Db 122 GTCTTCTCATCTACGAGG 141

RESULT 8
LOCUS AA629140
DEFINITION af60506.s1 Soares_NHMPu_S1 Homo sapiens cDNA clone IMAGE:1046387
(HUMAN); mRNA sequence.
ACCESSION AA629140
VERSION AA629140.1 GI:2541527
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 345)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisler,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Lennon,G., Marra,M., Martin
J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wyllie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: polyt not found
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 290.
Location/Qualifiers
1..345
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1046387"
/clone_lib="Soares_NHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/notes="Organ: mixed (see below); Vector: pT7n3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NBHPU, and fetal heart NBH19W) were mixed, and as circles
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

```



```

BASE COUNT      67 a   110 c   128 g   60 t
ORIGIN

Query Match      1.3%; Score 20; DB 9; Length 365;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 505 gtcttctcatctacgagg 524
Db 167 GTCTTCTCATCTACGAGG 186

RESULT 11
BM152194
LOCUS      365 bp mRNA linear EST 30-NOV-2001
DEFINITION TCBAp16167 Pediatric pre-B cell acute lymphoblastic leukemia
            Baylor-HGSC project-TCBA Homo sapiens cDNA clone TCBAp16167, mRNA
            sequence.
ACCESSION  BM152194
VERSION     BM152194.1 GI:17176285
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 365)
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
            Wei, Y., Tsang, Y. T. M., Mei, G., Ku, J. M., Ali-Osman, F. R. Jr.,
            Gunaratne, P. H., Muzny, D., Bouck, J., Gibbs, R. A. and Margolin, J. F.
            Pediatric Leukemia cDNA Sequencing Project (2001)
            Unpublished (2001)
            Contact: Dr. Judith F. Margolin
            Texas Children's Cancer Center and Human Genome Sequencing Center
            at Baylor College of Medicine
            1102 Bates, MC3-3320 Houston, TX 77030, USA
            Tel: 832-824-4536
            Fax: 832-825-4038
            Email: clones@ccc.org
            Seq primer: M13 primer.

FEATURES             Location/Qualifiers
     source           1..365
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="TCBAp167"
                     /clone_lib="Pediatric pre-B cell acute lymphoblastic
                     leukemia Baylor-HGSC project-TCBA"
                     /sex="male"
                     /tissue_type="leukopheresis"
                     /cell_type="pre-B cell"
                     /dev_stage="pediatric 2 years"
                     /lab_host="DH10B"
                     /note="Vector: lambda pSB; Site_1: BamHI; Site_2: EcoRI;
                     First strand cDNA was primed with an anchored
                     XhoI-oligo(dT) primer [5'GGAGGACTCGAGCGCGGAGGAG(T)VN
                     3'; V=A,C,G; N=A,C,G,T] and then dg tailed. Second strand
                     was primed with a BamHI-dC primer
                     [5'AGAGGTCGATCCGCGCGGCAATATTAAT(C) 3'].
                     Double-stranded cDNA was then digested with BamHI and XhoI
                     and directionally cloned into the BamHI and SalI sites of
                     lambda pSB vector. Library went through one round of
                     normalization. Library was constructed by Wei Yu at RIKEN
                     of Japan (Carninci P, Westover A, Nishiyama Y, Ohsumi T,
                     Itoh M, Nagaoka S, Sasakawa, Okazaki Y, Muramatsu M,
                     Schneider C, Hayashizaki Y, High efficiency selection of
                     full-length cDNA by improved biotinylated cap trapper..
                     DNA Res 4: 1, 61-6, Feb 28, 1997)"
BASE COUNT      66 a   111 c   126 g   62 t
ORIGIN

Query Match      1.3%; Score 20; DB 10; Length 365;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

BASE COUNT      67 a   110 c   128 g   60 t
ORIGIN

Query Match      1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 505 gtcttctcatctacgagg 524
Db 159 GTCTTCTCATCTACGAGG 178

RESULT 12
AW222767/c
LOCUS      391 bp mRNA linear EST 18-MAY-2001
DEFINITION AW222767 tomato fruit red ripe, TAMU Lycopersicon esculentum cDNA
            clone cLEN9E24, mRNA sequence.
ACCESSION  AW222767
VERSION     AW222767.1 GI:6534451
KEYWORDS    EST.
SOURCE      tomato.
ORGANISM    Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon
            1 (bases 1 to 391)
            Alcalá, J., Vrabalov, J., White, R., Matern, A. L., Holt, I. E., Liang, F.,
            Upton, J., Hansen, R., Craven, M. B., Bowman, C. L., Ann, S., Renning, C. M.,
            Fraser, C. M., Martin, G. B., Tankley, S. D. and Giovannoni, J.
            Generation of ESTs from tomato fruit tissue
            Unpublished (1999)
            Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            5 prime sequence.

FEATURES             Location/Qualifiers
     source           1..391
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLEN9E24"
                     /clone_lib="tomato fruit red ripe, TAMU"
                     /tissue_type="pericarp"
                     /dev_stage="red ripe (7-20 days post-breaker)"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; supplier: Giovannoni; Fruit were tagged at the
                     breaker stage (first sign of lycopene accumulation on the
                     blossom end of the fruit) and harvested 7 days
                     post-breaker (fully red-ripe), 10 days post-breaker, and
                     20 days post-breaker (over-ripe). 20 day fruit which
                     showed external or internal signs of pathogenesis were
                     discarded. Fruit were cut in half and the seeds and
                     locules were discarded prior to freezing the pericarp."
BASE COUNT      90 a   89 c   125 g   87 t
ORIGIN

Query Match      1.3%; Score 20; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 78 aggtgtgtcgtccgagtc 97
Db 351 AGGTGTGCTCGGAGTCCC 332

RESULT 13
AL120231
LOCUS      434 bp mRNA linear EST 25-FEB-2000
DEFINITION DKF20761G067.r1.761 (synonym: hamy2) Homo sapiens cDNA clone
            DKF20761G067 5', mRNA sequence.
ACCESSION  AL120231
VERSION     AL120231.1 GI:5926130
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 434)
Koehler, K., Beyer, A., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
EST (Koehler, et al.)
Unpublished (1999)
Contact: Koehler K
MIPS

Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center at the Charité, Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
No si sequence available.
This clone (DKFZp761G067) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

Source
1. 434
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp761G067"
/clone_lib="761 (synonym: hamy2)"
/tissue_type="amygdala"
/dev_stage="adult"
/lab_host="DH10B"
/note="vector: pSport1; Site_1: NotI; Site_2: SalI"

BASE COUNT
ORIGIN

77 a 129 c 157 g 71 t

Query Match 1.38; Score 20; DB 9; Length 434;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 505 gctctctcatctacgagg 524
|||||
Db 181 GCTCTCTCATCTACGAGG 200

RESULT 14
BM174982
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BM174982 455 bp mRNA linear EST 06-DEC-2001
TgESTzyb17c01.y1 TgRH Tachyzoite Subtracted cDNA Library Toxoplasma gondii cDNA clone TgESTzyb17c01.y1 5', mRNA sequence.
BM174982
BM174982.1 GI:17393200
EST.
Toxoplasma gondii.
Toxoplasma gondii
Eukaryota; Alveolata; Apicomplexa; Coccidia; Elmeriida;
Sarcocystidae; Toxoplasma.
1 (bases 1 to 455)
Tang, K., Cole, R., Fogarty, S., Sibley, L.D., Ajioka, J.A., White, M., Clifton, S., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagaris, R., Ronko, I., Kennedy, S., Maguire, L., Waterston, R. and Wilson, R.
Toxoplasma EST Project
Unpublished (2001)
Contact: Clifton, S.
Toxoplasma EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxowatson.wustl.edu
Contact David Sibley (toxowatson@wustl.edu) for further information relating to organism, libraries, or clone availability.
Seq primer: -40RP from Gibco
High quality sequence stop: 420.

FEATURES

Location/Qualifiers
1. 455
/organism="Toxoplasma gondii"
/strain="RH (Type I)"
/db_xref="taxon:5811"
/clone_lib="TgRH Tachyzoite Subtracted cDNA Library"
/dev_stage="Tachyzoite"
/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
/note="Vector: pBluscript SK-; Site_1: EcoRI; Site_2: XhoI
; Toxoplasma RH strain tachyzoites were grown in human foreskin fibroblast cultures in vitro. The library was originally constructed by K.L.Wan, Cambridge University. cDNAs were synthesized from polyA RNAs by oligo d(T) priming and directionally cloned into the EcoRI to XhoI sites of the Lambda ZapII vector using the ZAP-CDNA synthesis kit (Stratagene). The primary cDNA library was mass excised as phagemid using ExAssist helper phage (Stratagene). Phagemid DNA was extracted by phenol-chloroform method, and hybridized against a pool of over-represented ESTs (N>=12, from 5596 previous reads). The subtracted library was electroporated into DH10B (GeneHog, Invitrogen, Inc). WARNING: the library contains a small percentage of cDNAs derived from the human host cells. Library Source: David Sibley, Washington University."

BASE COUNT 87 a 122 c 154 g 92 t

ORIGIN

Query Match 1.38; Score 20; DB 10; Length 455;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1329 accgcgtgaatcggtttc 1348
|||||
Db 327 ACCGCGTGAATCGGTTTC 346

RESULT 15
BE241865
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BE241865 455 bp mRNA linear EST 03-OCT-2001
TCAAP2E0007 Pediatric acute myelogenous leukemia cell (FAB M1)
Baylor-HGSC project-TCAA Homo sapiens cDNA clone TCAAP0007, mRNA sequence.
BE241865
BE241865.1 GI:9093579
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 455)
Wei, Y., Tsang, Y.T.M., Mei, G., Ku, J.M., Ali-Osman Jr., F.R., Muzny, D., Bouck, J., Gibbs, R.A. and Margolin, J.F.
Pediatric Leukemia cDNA Sequencing Project
Unpublished (2000)
Contact: Dr. Judith F. Margolin
Texas Children's Cancer Center and Human Genome Sequencing Center
at Baylor College of Medicine
1102 Bates, MC3-3320 Houston, TX 77030, USA
Tel: 832-824-4536
Fax: 832-825-4038
Email: clones@txccc.org
Citation: Carninci, P. and Hayashizaki, Y. High efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Seq primer: M13 primer.

FEATURES

Location/Qualifiers
1. 455
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TCAAP0007"
/clone_lib="Pediatric acute myelogenous leukemia cell (FAB

```
M1) Baylor-HGSC project-TCAA"
/sex="male"
/tissue_type="leukopheresis"
/cell_type="myeloid cell"
/dev_stage="pediatric 6 years"
/lab_host="DH108"
/notes="Vector: lambda PSB; Site_1: BamHI; Site_2: EcoRI;
First strand cDNA was primed with an anchored
XhoI-oligo(dT) primer [5'GGAGGACTCGAGCGCGCAGGAG(T)VN
3'; V-A,C,G; N-A,C,G,T] and then dg tailed. Second strand
was primed with a BamHI-dC primer
[5'AGAGAGCTCGGATCGCGCGCAATAATAAT(C) 3'].
Double-stranded cDNA was then digested with BamHI and XhoI
and directionally cloned into the BamHI and SalI sites of
lambda PSB vector. Library went through one round of
normalization. Library was constructed by Wei Yu at RIKEN
of Japan (Carninci P, Westover A, Nishiyama Y, Ohsumi T,
Itoh M, Nagaoaka S, Sasaki N, Okazaki Y, Muramatsu M,
Schneider C, Hayashizaki Y, High efficiency selection of
full-length cDNA by improved biotinylated cap trapper.,
DNA Res 4: 1, 61-6, Feb 28, 1997)"
BASE COUNT      83 a   133 c   164 g   75 t
ORIGIN
```

```
Query Match      1.38; Score 20; DB 10; Length 455;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 505 gtcttctctatctacgaggg 524
    |||||||
Db 158 GTCTTCCTCATCTACGAGGG 177
```

Search completed: April 27, 2002, 02:50:50
Job time: 13442 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 07:44:44 ; Search time 95.74 Seconds
(without alignments)
3027.782 Million cell updates/sec

Title: US-09-882-694A-7
Perfect score: 1578
Sequence: 1 atgactccagaccagtg.....aattcaagttgggcatga 1578

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 139433 seqs, 91850399 residues

Word size : 0
Total number of hits satisfying chosen parameters: 278866

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Pending Patents NA-New.*
1: /cgn2.6/prodata/1/pna/PCT_NEW_COMB.seq.*
2: /cgn2.6/prodata/1/pna/US06_NEW_COMB.seq.*
3: /cgn2.6/prodata/1/pna/US07_NEW_COMB.seq.*
4: /cgn2.6/prodata/1/pna/US08_NEW_COMB.seq.*
5: /cgn2.6/prodata/1/pna/US09_NEW_COMB.seq.*
6: /cgn2.6/prodata/1/pna/US10_NEW_COMB.seq.*
7: /cgn2.6/prodata/1/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 17 | 1.1 | 1497 | 5 | US-09-540-209B-796 |
| 2 | 16 | 1.0 | 209 | 5 | US-09-540-210B-6207 |
| 3 | 16 | 1.0 | 238 | 5 | US-09-975-254-11574 |
| 4 | 16 | 1.0 | 239 | 5 | US-09-540-210B-927 |
| 5 | 16 | 1.0 | 254 | 5 | US-09-975-254-8433 |
| 6 | 16 | 1.0 | 254 | 5 | US-09-975-254-15910 |
| 7 | 16 | 1.0 | 260 | 5 | US-09-540-210B-27025 |
| 8 | 16 | 1.0 | 271 | 5 | US-09-975-254-25409 |
| 9 | 16 | 1.0 | 279 | 5 | US-09-540-210B-26631 |
| 10 | 16 | 1.0 | 292 | 5 | US-09-540-210B-35245 |
| 11 | 16 | 1.0 | 294 | 5 | US-09-789-189-1738 |
| 12 | 16 | 1.0 | 619 | 6 | US-10-103-313-212 |
| 13 | 16 | 1.0 | 840 | 5 | US-09-540-209B-250 |
| 14 | 16 | 1.0 | 857 | 7 | US-60-365-384-443 |
| 15 | 16 | 1.0 | 1004 | 7 | US-60-365-264-87 |
| 16 | 16 | 1.0 | 1428 | 1 | PCT-US02-07826-70 |
| 17 | 16 | 1.0 | 1428 | 6 | US-10-097-340-70 |
| 18 | 16 | 1.0 | 1504 | 6 | US-10-105-299-476 |
| 19 | 16 | 1.0 | 1972 | 6 | US-10-105-299-7237 |
| 20 | 16 | 1.0 | 1981 | 6 | US-10-105-299-467 |
| 21 | 16 | 1.0 | 1982 | 7 | US-60-365-384-96 |
| 22 | 16 | 1.0 | 2200 | 6 | US-10-105-299-2789 |
| 23 | 16 | 1.0 | 2623 | 6 | US-10-112-699-2907 |
| 24 | 16 | 1.0 | 2710 | 1 | PCT-US02-09921-518 |
| 25 | 16 | 1.0 | 2861 | 6 | US-10-106-698-264 |
| 26 | 16 | 1.0 | 3516 | 6 | US-10-105-299-10923 |

| | | | | | | |
|------|----|-----|-------|---|---------------------|--------------------|
| C 27 | 16 | 1.0 | 3673 | 1 | PCT-US02-09921-408 | Sequence 408, App |
| C 28 | 16 | 1.0 | 4281 | 6 | US-10-116-802-153 | Sequence 153, App |
| C 29 | 16 | 1.0 | 4713 | 1 | PCT-US02-09944-386 | Sequence 386, App |
| C 30 | 16 | 1.0 | 5185 | 7 | US-60-365-264-401 | Sequence 401, App |
| C 31 | 16 | 1.0 | 10367 | 6 | US-10-105-299-12588 | Sequence 12588, A |
| C 32 | 16 | 1.0 | 16347 | 6 | US-10-105-299-6563 | Sequence 6563, A |
| C 33 | 16 | 1.0 | 20295 | 6 | US-10-105-299-7351 | Sequence 7351, App |
| C 34 | 16 | 1.0 | 20295 | 6 | US-10-105-299-7354 | Sequence 7354, App |
| C 35 | 16 | 1.0 | 21010 | 6 | US-10-105-299-7338 | Sequence 7338, App |
| C 36 | 16 | 1.0 | 21010 | 6 | US-10-105-299-7338 | Sequence 7338, App |
| C 37 | 16 | 1.0 | 21024 | 6 | US-10-105-299-7339 | Sequence 7339, App |
| C 38 | 16 | 1.0 | 21024 | 6 | US-10-105-299-7339 | Sequence 7339, App |
| C 39 | 16 | 1.0 | 24413 | 1 | PCT-US02-09239-370 | Sequence 370, App |
| C 40 | 16 | 1.0 | 24413 | 6 | US-10-105-299-7346 | Sequence 7346, App |
| C 41 | 16 | 1.0 | 24413 | 6 | US-10-105-299-13916 | Sequence 13916, A |
| C 42 | 16 | 1.0 | 26439 | 6 | US-10-105-299-7958 | Sequence 7958, App |
| C 43 | 16 | 1.0 | 26439 | 6 | US-10-105-299-12665 | Sequence 12665, A |
| C 44 | 16 | 1.0 | 31328 | 6 | US-10-109-310-19 | Sequence 19, Appl |
| C 45 | 15 | 1.0 | 93 | 5 | US-09-975-254-10253 | Sequence 10253, A |

ALIGNMENTS

RESULT 1
US-09-540-209B-796
; Sequence 796, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES
; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540.209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 796
; LENGTH: 1497
; TYPE: DNA
; ORGANISM: B.fragilis
US-09-540-209B-796

Query Match 1.1%; Score 17; DB 5; Length 1497;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 771 cgcagcaatgcactctgg 787
|||||
Db 632 cgcagcaatgcactctgg 648

RESULT 2
US-09-540-210B-6207/c
; Sequence 6207, Application US/09540210B
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.
; APPLICANT: Deleane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullaney, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF URINARY TRACT TISSUE
; FILE REFERENCE: PD-1037 CIP
; CURRENT APPLICATION NUMBER: 2002-04-03
; PRIOR FILING DATE: November 18, 1997
; PRIOR APPLICATION NUMBER: 08/972,899
; PRIOR FILING DATE: February 27, 1995
; PRIOR APPLICATION NUMBER: 08/395,244
; PRIOR FILING DATE: September 27, 1996
; PRIOR APPLICATION NUMBER: 60/005,526
; PRIOR FILING DATE: September 29, 1995

; PRIOR APPLICATION NUMBER: 08/824,029
; PRIOR FILING DATE: March 25, 1997
; PRIOR APPLICATION NUMBER: 60/014,010
; PRIOR FILING DATE: March 25, 1996
; PRIOR APPLICATION NUMBER: 08/826,847
; PRIOR FILING DATE: April 10, 1997
; PRIOR APPLICATION NUMBER: 60/015,533
; PRIOR FILING DATE: April 10, 1996
; PRIOR APPLICATION NUMBER: 08/903,555
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/023,308
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/862,178
; PRIOR FILING DATE: May 22, 1997
; PRIOR APPLICATION NUMBER: 60/018,217
; PRIOR FILING DATE: May 23, 1996
; PRIOR APPLICATION NUMBER: 08/881,589
; PRIOR FILING DATE: June 24, 1997
; PRIOR APPLICATION NUMBER: 60/021,275
; PRIOR FILING DATE: June 25, 1996
; PRIOR APPLICATION NUMBER: 08/903,802
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/023,308
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/905,881
; PRIOR FILING DATE: August 1, 1997
; PRIOR APPLICATION NUMBER: 60/025,204
; PRIOR FILING DATE: August 1, 1996
; PRIOR APPLICATION NUMBER: 08/903,471
; PRIOR FILING DATE: July 30, 1997
; PRIOR APPLICATION NUMBER: 60/025,478
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/903,556
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/025,217
; PRIOR FILING DATE: August 22, 1996
; PRIOR APPLICATION NUMBER: 08/937,142
; PRIOR FILING DATE: September 23, 1997
; PRIOR APPLICATION NUMBER: 60/026,598
; PRIOR FILING DATE: September 24, 1996
; PRIOR APPLICATION NUMBER: 08/960,746
; PRIOR FILING DATE: October 29, 1997
; PRIOR APPLICATION NUMBER: 60/030,144
; PRIOR FILING DATE: October 30, 1996
; PRIOR APPLICATION NUMBER: 08/826,847
; PRIOR FILING DATE: April 10, 1997
; PRIOR APPLICATION NUMBER: 60/015,533
; PRIOR FILING DATE: April 10, 1996
; PRIOR APPLICATION NUMBER: 08/755,524
; PRIOR FILING DATE: November 22, 1996
; PRIOR APPLICATION NUMBER: 60/007,495
; PRIOR FILING DATE: November 22, 1995
; PRIOR APPLICATION NUMBER: 09/021,031
; PRIOR FILING DATE: February 10, 1998
; PRIOR APPLICATION NUMBER: 60/039,325
; PRIOR FILING DATE: February 13, 1997
; PRIOR APPLICATION NUMBER: 09/035,172
; PRIOR FILING DATE: March 4, 1998
; PRIOR APPLICATION NUMBER: 60/040,431
; PRIOR FILING DATE: March 5, 1997
; PRIOR APPLICATION NUMBER: 09/041,894
; PRIOR FILING DATE: March 12, 1998
; PRIOR APPLICATION NUMBER: 60/040,199
; PRIOR FILING DATE: March 14, 1997
; PRIOR APPLICATION NUMBER: 09/050,817
; PRIOR FILING DATE: March 30, 1998
; PRIOR APPLICATION NUMBER: 60/043,792
; PRIOR FILING DATE: April 11, 1997
; PRIOR APPLICATION NUMBER: 09/074,999
; PRIOR FILING DATE: May 8, 1998
; PRIOR APPLICATION NUMBER: 60/048,431
; PRIOR FILING DATE: May 29, 1997
; PRIOR APPLICATION NUMBER: 09/107,592

; PRIOR FILING DATE: June 30, 1998
; PRIOR APPLICATION NUMBER: 60/052,751
; PRIOR FILING DATE: July 1, 1997
; PRIOR APPLICATION NUMBER: 09/094,079
; PRIOR FILING DATE: June 9, 1998
; PRIOR APPLICATION NUMBER: 60/049,975
; PRIOR FILING DATE: June 13, 1997
; NUMBER OF SEQ ID NOS: 35654
; SOFTWARE: PERL Program
; SEQ ID NO 6207
; LENGTH: 209
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: hu00736504
; NAME/KEY: unsure
; LOCATION: 35, 71
; OTHER INFORMATION: a, t, c, g, or other
US-09-540-210B-6207

Query Match 1.0%; Score 16; DB 5; Length 209;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 478 cacacatactccagg 493
|||||
Db 203 CACACATACTCCAGG 188

RESULT 3

US-09-975-254-11574/c
; Sequence 11574, Application US/09975254
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Heck, Gregory R.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15309)B
; CURRENT APPLICATION NUMBER: US/09/975,254
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US/09/263,191
; PRIOR FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 31255
; SEQ ID NO 11574
; LENGTH: 238
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 700904990h1
US-09-975-254-11574

Query Match 1.0%; Score 16; DB 5; Length 238;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 66 gacggcgccgaggt 81
|||||
Db 158 GACGGCGCCGAGGT 143

RESULT 4

US-09-540-210B-927
; Sequence 927, Application US/09540210B
; GENERAL INFORMATION:
; APPLICANT: Sellhauer, Jeffrey J.
; APPLICANT: Deleageane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullany, Sara J.
; APPLICANT: Naughton, Rebecca E.

| FILE REFERENCE | PD-1037 CIP | TITLE OF INVENTION: POLYNUCLEOTIDES OF URINARY TRACT TISSUE |
|----------------|--|---|
| 1 | FILE REFERENCE: PD-1037 CIP | |
| 2 | CURRENT APPLICATION NUMBER: US/09/540,210B | |
| 3 | CURRENT FILING DATE: 2002-04-03 | |
| 4 | PRIOR APPLICATION NUMBER: 08/972,899 | |
| 5 | PRIOR FILING DATE: November 18, 1997 | |
| 6 | PRIOR APPLICATION NUMBER: 08/395,244 | |
| 7 | PRIOR FILING DATE: February 27, 1995 | |
| 8 | PRIOR APPLICATION NUMBER: 08/722,922 | |
| 9 | PRIOR FILING DATE: September 27, 1996 | |
| 10 | PRIOR APPLICATION NUMBER: 60/005,526 | |
| 11 | PRIOR FILING DATE: September 29, 1995 | |
| 12 | PRIOR APPLICATION NUMBER: 08/824,029 | |
| 13 | PRIOR FILING DATE: March 25, 1997 | |
| 14 | PRIOR APPLICATION NUMBER: 60/014,010 | |
| 15 | PRIOR FILING DATE: March 25, 1996 | |
| 16 | PRIOR APPLICATION NUMBER: 08/826,847 | |
| 17 | PRIOR FILING DATE: April 10, 1997 | |
| 18 | PRIOR APPLICATION NUMBER: 60/015,533 | |
| 19 | PRIOR FILING DATE: April 10, 1996 | |
| 20 | PRIOR APPLICATION NUMBER: 08/903,555 | |
| 21 | PRIOR FILING DATE: July 31, 1997 | |
| 22 | PRIOR APPLICATION NUMBER: 60/023,308 | |
| 23 | PRIOR FILING DATE: July 31, 1996 | |
| 24 | PRIOR APPLICATION NUMBER: 08/862,178 | |
| 25 | PRIOR FILING DATE: May 22, 1997 | |
| 26 | PRIOR APPLICATION NUMBER: 60/018,217 | |
| 27 | PRIOR FILING DATE: May 23, 1996 | |
| 28 | PRIOR APPLICATION NUMBER: 08/881,589 | |
| 29 | PRIOR FILING DATE: June 24, 1997 | |
| 30 | PRIOR APPLICATION NUMBER: 60/021,275 | |
| 31 | PRIOR FILING DATE: June 25, 1996 | |
| 32 | PRIOR APPLICATION NUMBER: 08/903,802 | |
| 33 | PRIOR FILING DATE: July 31, 1997 | |
| 34 | PRIOR APPLICATION NUMBER: 60/023,308 | |
| 35 | PRIOR FILING DATE: July 31, 1996 | |
| 36 | PRIOR APPLICATION NUMBER: 08/905,881 | |
| 37 | PRIOR FILING DATE: August 1, 1997 | |
| 38 | PRIOR APPLICATION NUMBER: 60/025,204 | |
| 39 | PRIOR FILING DATE: August 1, 1996 | |
| 40 | PRIOR APPLICATION NUMBER: 08/903,471 | |
| 41 | PRIOR FILING DATE: July 30, 1997 | |
| 42 | PRIOR APPLICATION NUMBER: 60/025,478 | |
| 43 | PRIOR FILING DATE: July 31, 1996 | |
| 44 | PRIOR APPLICATION NUMBER: 08/903,556 | |
| 45 | PRIOR FILING DATE: July 31, 1997 | |
| 46 | PRIOR APPLICATION NUMBER: 60/025,217 | |
| 47 | PRIOR FILING DATE: August 22, 1996 | |
| 48 | PRIOR APPLICATION NUMBER: 08/937,142 | |
| 49 | PRIOR FILING DATE: September 23, 1997 | |
| 50 | PRIOR APPLICATION NUMBER: 60/026,598 | |
| 51 | PRIOR FILING DATE: September 24, 1996 | |
| 52 | PRIOR APPLICATION NUMBER: 08/960,746 | |
| 53 | PRIOR FILING DATE: October 29, 1997 | |
| 54 | PRIOR APPLICATION NUMBER: 60/030,144 | |
| 55 | PRIOR FILING DATE: October 30, 1996 | |
| 56 | PRIOR APPLICATION NUMBER: 08/826,847 | |
| 57 | PRIOR FILING DATE: April 10, 1997 | |
| 58 | PRIOR APPLICATION NUMBER: 60/015,533 | |
| 59 | PRIOR FILING DATE: April 10, 1996 | |
| 60 | PRIOR APPLICATION NUMBER: 08/755,524 | |
| 61 | PRIOR FILING DATE: November 22, 1996 | |
| 62 | PRIOR APPLICATION NUMBER: 60/007,495 | |
| 63 | PRIOR FILING DATE: November 22, 1995 | |
| 64 | PRIOR APPLICATION NUMBER: 09/021,031 | |
| 65 | PRIOR FILING DATE: February 10, 1998 | |
| 66 | PRIOR APPLICATION NUMBER: 60/039,325 | |
| 67 | PRIOR FILING DATE: February 13, 1997 | |
| 68 | PRIOR APPLICATION NUMBER: 09/035,172 | |
| 69 | PRIOR FILING DATE: March 4, 1998 | |
| 70 | PRIOR APPLICATION NUMBER: 60/040,431 | |
| 71 | PRIOR FILING DATE: March 5, 1997 | |
| 72 | PRIOR APPLICATION NUMBER: 09/041,894 | |

```

> PRIOR FILING DATE: March 12, 1998
> PRIOR APPLICATION NUMBER: 60/040,199
> PRIOR FILING DATE: March 14, 1997/040,199
> PRIOR APPLICATION NUMBER: 60/050,817
> PRIOR FILING DATE: March 30, 1998
> PRIOR APPLICATION NUMBER: 60/043,792
> PRIOR FILING DATE: April 11, 1997
> PRIOR APPLICATION NUMBER: 60/074,999
> PRIOR FILING DATE: May 8, 1998
> PRIOR APPLICATION NUMBER: 60/048,431
> PRIOR FILING DATE: May 29, 1997
> PRIOR APPLICATION NUMBER: 60/107,592
> PRIOR FILING DATE: June 30, 1998
> PRIOR APPLICATION NUMBER: 60/052,751
> PRIOR FILING DATE: July 1, 1997
> PRIOR APPLICATION NUMBER: 60/094,079
> PRIOR FILING DATE: June 9, 1998
> PRIOR APPLICATION NUMBER: 60/049,975
> PRIOR FILING DATE: June 13, 1997
> NUMBER OF SEQ ID NOS: 35654
> SOFTWARE: PERL Program
> SEQ ID NO 927
> LENGTH: 239
> TYPE: DNA
> ORGANISM: Homo sapiens
> FEATURE:
> NAME/KEY: misc_feature
> OTHER INFORMATION: Incyte ID No.: hu00948023
> US-09-4108-927

```

Query Match 1.0%; Score 16; DB 5; Length 239;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 16: Conservative 0: Mismatches 0: Indels

Qy 184 tcaggcagttctccagc 199
 |||||
 Db 19 tcaggcagttctccagc 34

```

RESULT          5
> US-09-975-254-8433/C
> SEQUENCE 8433: Application US/09975254
> GENERAL INFORMATION:
> APPLICANT: Byrum, Joseph R.
> APPLICANT: Heck, Gregory R.
> APPLICANT: La Rosa, Thomas J.
> TITLE OF INVENTION: Nucleic Acid Molecule
> TITLE OF INVENTION: Plants
> FILE REFERENCE: 38-21(113509)B
> CURRENT APPLICATION NUMBER: US/09/975,254
> CURRENT FILING DATE: 2001-10-12
> PRIOR APPLICATION NUMBER: US/09/263,191
> PRIOR FILING DATE: 1999-03-05
> NUMBER OF SEQ ID NOS: 31255
> SEQ ID NO 8433
> LENGTH: 254
> TYPE: DNA
> ORGANISM: Glycine max
> OTHER INFORMATION: Clone ID: 700754577H1
US-09-975-254-8433

```

Query Match 1.0%; Score 16; DB 5; Length 254;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 16: Conservative 0; Mismatches 0; Indels

Qy 66 gacggcggcgaggt 81
|||||
Db 159 GACGGCGCGGCAGGT 144

RESULT 6

```
US-09-975-254-15910/c
; Sequence 15910, Application US/09975254
; GENERAL INFORMATION:
; APPLICANT: Byrum, Gregory R.
; APPLICANT: Heck, Gregory R.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15309)B
; CURRENT APPLICATION NUMBER: US/09/975,254
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US/09/263,191
; PRIOR FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 31255
; SEQ ID NO 15910
; LENGTH: 254
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 700952522H1
US-09-975-254-15910

Query Match          1.0%  Score 16; DB 5; Length 254;
Best Local Similarity 100.0%  Pred No. 58;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 gacgagcggcggcaggt 81
Db 78 GACGCGCGCGCAGGT 63

RESULT 7
US-09-540-210B-27025/c
; Sequence 27025, Application US/09540210B
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF URINARY TRACT TISSUE
; FILE REFERENCE: PD-1037 CIP
; CURRENT APPLICATION NUMBER: US/09/540,210B
; CURRENT FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 08/972,899
; PRIOR FILING DATE: November 18, 1997
; PRIOR APPLICATION NUMBER: 08/395,244
; PRIOR FILING DATE: February 27, 1995
; PRIOR APPLICATION NUMBER: 08/722,922
; PRIOR FILING DATE: September 27, 1996
; PRIOR APPLICATION NUMBER: 60/005,526
; PRIOR FILING DATE: September 29, 1995
; PRIOR APPLICATION NUMBER: 08/824,029
; PRIOR FILING DATE: March 25, 1997
; PRIOR APPLICATION NUMBER: 60/014,010
; PRIOR FILING DATE: March 25, 1996
; PRIOR APPLICATION NUMBER: 08/826,847
; PRIOR FILING DATE: April 10, 1997
; PRIOR APPLICATION NUMBER: 60/015,533
; PRIOR FILING DATE: April 10, 1996
; PRIOR APPLICATION NUMBER: 08/303,555
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/023,308
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/862,178
; PRIOR FILING DATE: May 22, 1997
; PRIOR APPLICATION NUMBER: 60/018,217
; PRIOR FILING DATE: May 23, 1996
; PRIOR APPLICATION NUMBER: 08/881,589
; PRIOR FILING DATE: June 24, 1997
; PRIOR APPLICATION NUMBER: 60/021,275
; PRIOR FILING DATE: June 25, 1996

; PRIOR APPLICATION NUMBER: 08/903,802
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/023,308
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/905,881
; PRIOR FILING DATE: August 1, 1997
; PRIOR APPLICATION NUMBER: 60/025,204
; PRIOR FILING DATE: August 1, 1996
; PRIOR APPLICATION NUMBER: 08/903,471
; PRIOR FILING DATE: July 30, 1997
; PRIOR APPLICATION NUMBER: 60/025,478
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/903,556
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/025,217
; PRIOR FILING DATE: August 22, 1996
; PRIOR APPLICATION NUMBER: 08/937,142
; PRIOR FILING DATE: September 23, 1997
; PRIOR APPLICATION NUMBER: 60/026,598
; PRIOR FILING DATE: September 24, 1996
; PRIOR APPLICATION NUMBER: 08/960,746
; PRIOR FILING DATE: October 29, 1997
; PRIOR APPLICATION NUMBER: 60/030,144
; PRIOR FILING DATE: October 30, 1996
; PRIOR APPLICATION NUMBER: 08/826,847
; PRIOR FILING DATE: April 10, 1997
; PRIOR APPLICATION NUMBER: 60/015,533
; PRIOR FILING DATE: April 10, 1996
; PRIOR APPLICATION NUMBER: 08/755,524
; PRIOR FILING DATE: November 22, 1996
; PRIOR APPLICATION NUMBER: 60/007,495
; PRIOR FILING DATE: November 22, 1995
; PRIOR APPLICATION NUMBER: 09/021,031
; PRIOR FILING DATE: February 10, 1998
; PRIOR APPLICATION NUMBER: 60/039,325
; PRIOR FILING DATE: February 13, 1997
; PRIOR APPLICATION NUMBER: 09/035,172
; PRIOR FILING DATE: March 4, 1998
; PRIOR APPLICATION NUMBER: 60/040,431
; PRIOR FILING DATE: March 5, 1997
; PRIOR APPLICATION NUMBER: 09/041,894
; PRIOR FILING DATE: March 12, 1998
; PRIOR APPLICATION NUMBER: 60/040,199
; PRIOR FILING DATE: March 14, 1997
; PRIOR APPLICATION NUMBER: 09/050,817
; PRIOR FILING DATE: March 30, 1998
; PRIOR APPLICATION NUMBER: 60/043,792
; PRIOR FILING DATE: April 11, 1997
; PRIOR APPLICATION NUMBER: 09/074,999
; PRIOR FILING DATE: May 8, 1998
; PRIOR APPLICATION NUMBER: 60/048,431
; PRIOR FILING DATE: May 29, 1997
; PRIOR APPLICATION NUMBER: 09/107,592
; PRIOR FILING DATE: June 30, 1998
; PRIOR APPLICATION NUMBER: 60/052,751
; PRIOR FILING DATE: July 1, 1997
; PRIOR APPLICATION NUMBER: 09/094,079
; PRIOR FILING DATE: June 9, 1998
; PRIOR APPLICATION NUMBER: 60/049,975
; PRIOR FILING DATE: June 13, 1997
; NUMBER OF SEQ ID NOS: 35654
; SOFTWARE: PERL Program
; SEQ ID NO 27025
; LENGTH: 260
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: hu00786148
; NAME/KEY: unsure
; LOCATION: 31, 136
; OTHER INFORMATION: a, t, c, g, or other
US-09-540-210B-27025
```


Query Match 1.0%; Score 16; DB 5; Length 260;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1133 ctgttggtctctatt 1148
Db 57 CTGTTGCTCTATT 42

RESULT 8

US-09-975-254-25409
; Sequence 25409, Application US/09975254
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Heck, Gregory R.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15309)B
; CURRENT APPLICATION NUMBER: US/09/975,254
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US/09/263,191
; PRIOR FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 31255
; SEQ ID NO 25409
; LENGTH: 271
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 700964702H1
US-09-975-254-25409

Query Match 1.0%; Score 16; DB 5; Length 271;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1246 ctctctctctgaccc 1261
Db 154 ctctctctctgaccc 169

RESULT 9

US-09-540-2108-26631
; Sequence 26631, Application US/095402108
; GENERAL INFORMATION:
; APPLICANT: Selthamer, Jeffrey J.
; APPLICANT: Deleagane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullony, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF URINARY TRACT TISSUE
; FILE REFERENCE: PD-1037 CIP
; CURRENT APPLICATION NUMBER: US/09/540,2108
; CURRENT FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 08/972,899
; PRIOR FILING DATE: November 18, 1997
; PRIOR APPLICATION NUMBER: 08/395,244
; PRIOR FILING DATE: February 27, 1995
; PRIOR APPLICATION NUMBER: 08/722,922
; PRIOR FILING DATE: September 27, 1996
; PRIOR APPLICATION NUMBER: 60/005,526
; PRIOR FILING DATE: September 29, 1995
; PRIOR APPLICATION NUMBER: 08/824,029
; PRIOR FILING DATE: March 25, 1997
; PRIOR APPLICATION NUMBER: 60/014,010
; PRIOR FILING DATE: March 25, 1996
; PRIOR APPLICATION NUMBER: 08/826,847
; PRIOR FILING DATE: April 10, 1997
; PRIOR APPLICATION NUMBER: 60/015,533
; PRIOR FILING DATE: April 10, 1996
; NUMBER OF SEQ ID NOS: 35654

```
; SOFTWARE: PERL Program
; SEQ ID NO 26631
; LENGTH: 279
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: hu00027146
US-09-540-210B-26631

Query Match          1.0%; Score 16; DB 5; Length 279;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1086 ggtattttccactcat 1101
Db 253 ggtattttccactcat 268
|||||

RESULT 10
US-09-540-210B-35245/c
; Sequence 35245, Application US/09540210B
; GENERAL INFORMATION:
; APPLICANT: Sellhamer, Jeffrey J.
; APPLICANT: Deleageane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF URINARY TRACT TISSUE
; FILE REFERENCE: PD-1037 CIP
; CURRENT APPLICATION NUMBER: US/09/540,210B
; PRIOR FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 08/972,899
; PRIOR FILING DATE: November 18, 1997
; PRIOR APPLICATION NUMBER: 08/395,244
; PRIOR FILING DATE: February 27, 1995
; PRIOR APPLICATION NUMBER: 08/722,922
; PRIOR FILING DATE: September 27, 1996
; PRIOR APPLICATION NUMBER: 60/005,526
; PRIOR FILING DATE: September 29, 1995
; PRIOR APPLICATION NUMBER: 08/824,029
; PRIOR FILING DATE: March 25, 1997
; PRIOR APPLICATION NUMBER: 60/014,010
; PRIOR FILING DATE: March 25, 1996
; PRIOR APPLICATION NUMBER: 08/826,847
; PRIOR FILING DATE: April 10, 1997
; PRIOR APPLICATION NUMBER: 60/015,533
; PRIOR FILING DATE: April 10, 1996
; PRIOR APPLICATION NUMBER: 08/903,555
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/023,308
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/862,178
; PRIOR FILING DATE: May 22, 1997
; PRIOR APPLICATION NUMBER: 60/018,217
; PRIOR FILING DATE: May 23, 1996
; PRIOR APPLICATION NUMBER: 08/881,589
; PRIOR FILING DATE: June 24, 1997
; PRIOR APPLICATION NUMBER: 60/021,275
; PRIOR FILING DATE: June 25, 1996
; PRIOR APPLICATION NUMBER: 08/903,802
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/023,308
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/905,881
; PRIOR FILING DATE: August 1, 1997
; PRIOR APPLICATION NUMBER: 60/025,204
; PRIOR FILING DATE: August 1, 1996
; PRIOR APPLICATION NUMBER: 08/903,471
; PRIOR FILING DATE: July 30, 1997
; PRIOR APPLICATION NUMBER: 60/025,478
```

```
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/903,556
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/025,217
; PRIOR FILING DATE: August 22, 1996
; PRIOR APPLICATION NUMBER: 08/937,142
; PRIOR FILING DATE: September 23, 1997
; PRIOR APPLICATION NUMBER: 60/026,598
; PRIOR FILING DATE: September 24, 1996
; PRIOR APPLICATION NUMBER: 08/960,746
; PRIOR FILING DATE: October 29, 1997
; PRIOR APPLICATION NUMBER: 60/030,144
; PRIOR FILING DATE: October 30, 1996
; PRIOR APPLICATION NUMBER: 08/826,847
; PRIOR FILING DATE: April 10, 1997
; PRIOR APPLICATION NUMBER: 60/015,533
; PRIOR FILING DATE: April 10, 1996
; PRIOR APPLICATION NUMBER: 08/755,524
; PRIOR FILING DATE: November 22, 1996
; PRIOR APPLICATION NUMBER: 60/007,495
; PRIOR FILING DATE: November 22, 1995
; PRIOR APPLICATION NUMBER: 09/021,031
; PRIOR FILING DATE: February 10, 1998
; PRIOR APPLICATION NUMBER: 60/039,325
; PRIOR FILING DATE: February 13, 1997
; PRIOR APPLICATION NUMBER: 09/035,172
; PRIOR FILING DATE: March 4, 1998
; PRIOR APPLICATION NUMBER: 60/040,431
; PRIOR FILING DATE: March 5, 1997
; PRIOR APPLICATION NUMBER: 09/041,894
; PRIOR FILING DATE: March 12, 1998
; PRIOR APPLICATION NUMBER: 60/040,199
; PRIOR FILING DATE: March 14, 1997
; PRIOR APPLICATION NUMBER: 09/050,817
; PRIOR FILING DATE: March 30, 1998
; PRIOR APPLICATION NUMBER: 60/043,792
; PRIOR FILING DATE: April 11, 1997
; PRIOR APPLICATION NUMBER: 09/074,999
; PRIOR FILING DATE: May 8, 1998
; PRIOR APPLICATION NUMBER: 60/048,431
; PRIOR FILING DATE: May 29, 1997
; PRIOR APPLICATION NUMBER: 09/107,592
; PRIOR FILING DATE: June 30, 1998
; PRIOR APPLICATION NUMBER: 60/052,751
; PRIOR FILING DATE: July 1, 1997
; PRIOR APPLICATION NUMBER: 09/094,079
; PRIOR FILING DATE: June 9, 1998
; PRIOR APPLICATION NUMBER: 60/049,975
; PRIOR FILING DATE: June 13, 1997
; NUMBER OF SEQ ID NOS: 35654
; SOFTWARE: PERL Program
; SEQ ID NO 35245
; LENGTH: 292
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: rat00111322
US-09-540-210B-35245
```

```
Query Match          1.0%; Score 16; DB 5; Length 292;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 71 cgccggcagggtgtgc 86
Db 76 CGCCGCGCAGGTGTGC 61
|||||
```

```
RESULT 11
US-09-789-189-1738
; Sequence 1738, Application US/09789189
```

; GENERAL INFORMATION:
; APPLICANT: Lelias, Jean-Michel
; TITLE OF INVENTION: Human Polynucleotides and Polypeptides
; FILE REFERENCE: 25436/1720
; CURRENT APPLICATION NUMBER: US/09/789,189
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/183452
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 2005
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1738
; LENGTH: 294
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-789-189-1738

Query Match 1.0%; Score 16; DB 5; Length 294;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 348 accaaatcaatcaat 363
|||||

Db 1 accaaatcaatcaat 16

RESULT 12
US-10-103-313-212
; Sequence 212, Application US/10103313
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PZ07C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 212
; LENGTH: 619
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (610)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (618)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-103-313-212

Query Match 1.0%; Score 16; DB 6; Length 619;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1397 tcagcttccactgac 1412
|||||

Db 352 tcagcttccactgac 367

RESULT 13
US-09-540-209B-250/c
; Sequence 250, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FRP
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 250

; LENGTH: 840
; TYPE: DNA
; ORGANISM: B. fragilis
US-09-540-209B-250

Query Match 1.0%; Score 16; DB 5; Length 840;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1130 ttctgtttgtctct 1145
|||||

Db 522 TTCCTGTTGGTCTCT 507

RESULT 14
US-60-365-384-443/c
; Sequence 443, Application US/60365384
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Jian-rui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Zhang, Jie
; APPLICANT: Zhou, Ping
; APPLICANT: Wehrman, Tom
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Ghosh, Malabika
; APPLICANT: Zhao, Qing A.
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Xue, Aidong J.
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Weng, Gezhi
; TITLE OF INVENTION: Novel Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/60/365,384
; CURRENT FILING DATE: 2002-03-14
; NUMBER OF SEQ ID NOS: 666
; SOFTWARE: PT-FL-genes Version 6.0
; SEQ ID NO 443
; LENGTH: 857
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-365-384-443

Query Match 1.0%; Score 16; DB 7; Length 857;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 488 ccaggattcgtggca 503
|||||

Db 465 CCCAGGATTCGTGGCA 450

RESULT 15
US-60-365-264-87/c
; Sequence 87, Application US/60365264
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Jian-rui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Ghosh, Malabika
; APPLICANT: Xue, Aidong J.
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhao, Qing A.

; APPLICANT: Wang, Dunrui
; APPLICANT: Goodrich, Ryle W.
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wehrman, Tom
; APPLICANT: Meng, Gezhi
; TITLE OF INVENTION: Novel Nucleic Acids and
; FILE REFERENCE: 816
; CURRENT APPLICATION NUMBER: US/60/365,264
; CURRENT FILING DATE: 2002-03-14
; NUMBER OF SEQ ID NOS: 728
; SOFTWARE: pt_FL_genes Version 6.0
; SEQ ID NO: 87
; LENGTH: 1004
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (243)..(1004)
US-60-365-264-87

Query Match 1.0%; Score 16; DB 7; Length 1004;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 502 catgtcttctctctct 517
Db 530 CATGTCTTCTCTCATCT 515

Search completed: April 27, 2002, 07:44:48
Job time: 17140 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 07:42:26 ; Search time 8172.53 Seconds
(without alignments)
4177.393 Million cell updates/sec

Title: US-09-882-694A-7
Perfect score: 1578
Sequence: 1 atggactcagaccagtg.....aatttcagttgggccatga 1578

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2179536 seqs, 1081749327 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4359072

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending_Patents_NA_Main:*

- 1: /cgn2_6/ptodata/2/pna/pctus_COMB.seq:*
- 2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
- 3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
- 4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
- 5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
- 6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
- 7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
- 8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
- 9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
- 10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
- 11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
- 12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
- 13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
- 14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
- 15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
- 16: /cgn2_6/ptodata/2/pna/US092_COMB.seq:*
- 17: /cgn2_6/ptodata/2/pna/US093_COMB.seq:*
- 18: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
- 19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
- 20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
- 21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
- 22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
- 23: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
- 24: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
- 25: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
- 26: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
- 27: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
- 28: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
- 29: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
- 30: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
- 31: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
- 32: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
- 33: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
- 34: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
- 35: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
- 36: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
- 37: /cgn2_6/ptodata/2/pna/US100_COMB.seq:*
- 38: /cgn2_6/ptodata/2/pna/US101_COMB.seq:*
- 39: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
- 40: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
- 41: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
- 42: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
- 43: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*

- 44: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
- 45: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
- 46: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
- 47: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
- 48: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
- 49: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
- 50: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
- 51: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
- 52: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
- 53: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
- 54: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
- 55: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
- 56: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
- 57: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
- 58: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
- 59: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
- 60: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
- 61: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
- 62: /cgn2_6/ptodata/2/pna/US6023_COMB.seq:*
- 63: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
- 64: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
- 65: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
- 66: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
- 67: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
- 68: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
- 69: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
- 70: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
- 71: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
- 72: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
- 73: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
- 74: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
- 75: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 1578 | 100.0 | 1578 | 17 | US-09-351-224E-7 |
| 2 | 1578 | 100.0 | 1578 | 26 | US-09-677-488A-7 |
| 3 | 1578 | 100.0 | 1578 | 26 | US-09-677-682A-7 |
| 4 | 1578 | 100.0 | 1578 | 26 | US-09-677-682B-7 |
| 5 | 1578 | 100.0 | 1578 | 33 | US-09-882-694A-7 |
| 6 | 1578 | 100.0 | 1578 | 33 | US-09-882-694A-7 |
| 7 | 1578 | 100.0 | 1764 | 17 | US-09-351-224E-6 |
| 8 | 1578 | 100.0 | 1764 | 26 | US-09-677-488A-6 |
| 9 | 1578 | 100.0 | 1764 | 26 | US-09-677-682A-6 |
| 10 | 1578 | 100.0 | 1764 | 26 | US-09-677-682B-6 |
| 11 | 1578 | 100.0 | 1764 | 33 | US-09-882-694A-6 |
| 12 | 1578 | 100.0 | 1764 | 33 | US-09-882-694A-6 |
| 13 | 937 | 59.4 | 1578 | 17 | US-09-351-224-7 |
| 14 | 937 | 59.4 | 1578 | 26 | US-09-677-488-7 |
| 15 | 937 | 59.4 | 1578 | 26 | US-09-677-682-7 |
| 16 | 937 | 59.4 | 1764 | 17 | US-09-351-224-6 |
| 17 | 937 | 59.4 | 1764 | 26 | US-09-677-488-6 |
| 18 | 937 | 59.4 | 1764 | 26 | US-09-677-682-6 |
| 19 | 937 | 59.4 | 1764 | 26 | US-09-677-682-6 |
| 20 | 937 | 59.4 | 1764 | 26 | US-09-677-682-6 |
| C | | | | | |
| 21 | 21 | 1.3 | 29713 | 62 | US-60-230-435-897 |
| 22 | 21 | 1.3 | 32768 | 59 | US-60-207-315-186 |
| 23 | 21 | 1.3 | 38059 | 1 | PCT-US01-12012-1 |
| 24 | 21 | 1.3 | 38059 | 17 | US-09-328-925-4 |
| 25 | 21 | 1.3 | 38059 | 33 | US-09-680-107-2125 |
| 26 | 21 | 1.3 | 38059 | 37 | US-10-032-084-4 |
| 27 | 21 | 1.3 | 38059 | 73 | US-60-343-185-102 |
| 28 | 21 | 1.3 | 166 | 28 | US-09-704-424-5306 |
| 29 | 20 | 1.3 | 221 | 14 | US-09-016-866-1007 |
| 30 | 20 | 1.3 | 221 | 20 | US-09-532-315-12249 |
| 31 | 20 | 1.3 | 221 | 20 | US-09-532-315-12249 |

| | | | |
|----|------|--|------|
| Qy | 1441 | tacaacagtcgcattatagcgcttgcacttgctcttggtctcttgaaactgggtcgctgcat | 1500 |
| Db | 1441 | tacaacagtcgcattatagcgcttgcacttgctcttggtctcttgaaactgggtcgctgcat | 1500 |
| Qy | 1501 | gccaggaagcattatcaggagaccccaacttgagcttgacgacgggtcgctcgagagcagaa | 1560 |
| Db | 1501 | gccaggaagcattatcaggagaccccaacttgagcttgacgacgggtcgctcgagagcagaa | 1560 |
| Qy | 1561 | tttcaagttggggccatga | 1578 |
| Db | 1561 | tttcaagttggggccatga | 1578 |

```
Query Match      100.0%; Score 1578; DB 26; Length 1578;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

| | | | |
|----|------|--|------|
| Qy | 421 | agctcagtgaccatgatacctgctaacaagatcccgcgcgctgatagcgcctatagtcac | 480 |
| Db | 421 | | |
| Qy | 421 | agctcagtgaccatgatacctgctaacaagatcccgcgcgctgatagcgcctatagtcac | 480 |
| Db | 421 | | |
| Qy | 481 | acatactcccaggatcgtggatgctcttcctcatctacgaggagtcgcgctggtggtg | 540 |
| Db | 481 | | |
| Qy | 481 | acatactcccaggatcgtggatgctcttcctcatctacgaggagtcgcgctggtggtg | 540 |
| Db | 481 | | |
| Qy | 541 | ctcttggtcaactgtttgccttgaaaaaagaaaccttgggttcattgaaatcggattcggc | 600 |
| Db | 541 | | |
| Qy | 541 | ctcttggtcaactgtttgccttgaaaaaagaaaccttgggttcattgaaatcggattcggc | 600 |
| Db | 541 | | |
| Qy | 601 | ctcacgatcgtctctctggtatctctcttatacgccattctagcgcggtccaaacccaag | 660 |
| Db | 601 | | |
| Qy | 601 | ctcacgatcgtctctctggtatctctcttatacgccattctagcgcggtccaaacccaag | 660 |
| Db | 601 | | |
| Qy | 661 | gctccaaactcacaggatacgactgcttggaagcaactatactgctggtcgcagcgcgtc | 720 |
| Db | 661 | | |
| Qy | 661 | gctccaaactcacaggatacgactgcttggaagcaactatactgctggtcgcagcgcgtc | 720 |
| Db | 661 | | |
| Qy | 721 | tgcttaactcgtggcctttcgacatcctgcttcatgttcatgttggtcgtgacgcagcaatg | 780 |
| Db | 721 | | |
| Qy | 721 | tgcttaactcgtggcctttcgacatcctgcttcatgttcatgttggtcgtgacgcagcaatg | 780 |
| Db | 721 | | |
| Qy | 781 | catctggctgaagaatgcacagatgctgctctgatacgggtacccaaagcagtggtcagtcga | 840 |
| Db | 781 | | |
| Qy | 781 | catctggctgaagaatgcacagatgctgctctgatacgggtacccaaagcagtggtcagtcga | 840 |
| Db | 781 | | |
| Qy | 841 | atcataattggctcttgacccgcttccatatatacaatcgacgttctgtatggaaattaca | 900 |
| Db | 841 | | |
| Qy | 841 | atcataattggctcttgacccgcttccatatatacaatcgacgttctgtatggaaattaca | 900 |
| Db | 841 | | |
| Qy | 901 | gattcgcgactctattctaagttccgcggctctatttccattcgcagacaataacacgaatcc | 960 |
| Db | 901 | | |
| Qy | 901 | gattcgcgactctattctaagttccgcggctctatttccattcgcagacaataacacgaatcc | 960 |
| Db | 901 | | |
| Qy | 961 | cttcggtcgctcagttttgcaacggtccctctcatgtggcgtatcgtgatggcctctctc | 1020 |
| Db | 961 | | |
| Qy | 961 | cttcggtcgctcagttttgcaacggtccctctcatgtggcgtatcgtgatggcctctctc | 1020 |
| Db | 961 | | |
| Qy | 1021 | gcctcaaacgtctacaagagatcgctcgcactcactgcagcttggtccggcgacaat | 1080 |
| Db | 1021 | | |
| Qy | 1021 | gcctcaaacgtctacaagagatcgctcgcactcactgcagcttggtccggcgacaat | 1080 |
| Db | 1021 | | |
| Qy | 1081 | ggcgtggtattttccactcatctcgaaacgattcatcccgctggcgaagtctcgtttgg | 1140 |
| Db | 1081 | | |
| Qy | 1081 | ggcgtggtattttccactcatctcgaaacgattcatcccgctggcgaagtctcgtttgg | 1140 |
| Db | 1081 | | |
| Qy | 1141 | ttcttattcgcgacctgggaattcttgccacatcgcatatatttctaggttctcagc | 1200 |
| Db | 1141 | | |
| Qy | 1141 | ttcttattcgcgacctgggaattcttgccacatcgcatatatttctaggttctcagc | 1200 |
| Db | 1141 | | |
| Qy | 1201 | acagctttccaatgccttggttccaattccgcgcttgctaccagacaactctctctcgtatc | 1260 |
| Db | 1201 | | |
| Qy | 1201 | acagctttccaatgccttggttccaattccgcgcttgctaccagacaactctctctcgtatc | 1260 |
| Db | 1201 | | |
| Qy | 1261 | ccaatgcgcctactctctacaaaagcagatccaaagtctctgcagacactcgtgct | 1320 |
| Db | 1261 | | |
| Qy | 1261 | ccaatgcgcctactctctacaaaagcagatccaaagtctctgcagacactcgtgct | 1320 |
| Db | 1261 | | |
| Qy | 1321 | tttgtgtaccgcgctggaatcgggtttcttggtccaattgctctagcgggtggtcttcacgtcc | 1380 |
| Db | 1321 | | |
| Qy | 1321 | tttgtgtaccgcgctggaatcgggtttcttggtccaattgctctagcgggtggtcttcacgtcc | 1380 |
| Db | 1321 | | |
| Qy | 1381 | gtcacacagtgtttttcagcttccactgcacgtgcctacggcgcgtcaaccatgaat | 1440 |
| Db | 1381 | | |
| Qy | 1381 | gtcacacagtgtttttcagcttccactgcacgtgcctacggcgcgtcaaccatgaat | 1440 |
| Db | 1381 | | |
| Qy | 1441 | tacacaagtgcgattatagcgttgacgttgccttgccttgccttgaaactgggtcgatc | 1500 |
| Db | 1441 | | |
| Qy | 1441 | tacacaagtgcgattatagcgttgacgttgccttgccttgccttgaaactgggtcgatc | 1500 |
| Db | 1441 | | |

QY 1501 gccaggaagcattatcagggagccaccacttgagcttgagcggggtcgtcggagcagaa 1560
|||||
Db 1501 gccaggaagcattatcagggagccaccacttgagcttgagcggggtcgtcggagcagaa 1560
QY 1561 ttccaagttgagggccatga 1578
|||||
Db 1561 ttccaagttgagggccatga 1578
|||||

RESULT 3
US-09-677-682A-7
: Sequence 7, Application US/09677682A
: GENERAL INFORMATION:
: APPLICANT: Duwick, Jon
: APPLICANT: Maddox, Joyce
: APPLICANT: Gilliam, Jacob
: APPLICANT: Folkerts, Otto
: APPLICANT: Crasta, Oswald R.
: TITLE OF INVENTION: Compositions and Methods for Fumonisin
: TITLE OF INVENTION: Detoxification
: FILE REFERENCE: 35718/204101
: CURRENT APPLICATION NUMBER: US/09/677,682A
: PRIOR FILING DATE: 2000-10-02
: PRIOR APPLICATION NUMBER: 09/351,224
: PRIOR FILING DATE: 1999-07-12
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 7
: LENGTH: 1578
: TYPE: DNA
: ORGANISM: Exophiala spinifera
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: (0)-(0)
: OTHER INFORMATION: permealase, fully spliced cDNA
US-09-677-682A-7

Query Match 100.0%; Score 1578; DB 26; Length 1578;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggaactcagacaaagtggatagcggcagagaaaggcggagcagacacacgaagaac 60
|||||
Db 1 atggaactcagacaaagtggatagcggcagagaaaggcggagcagacacacgaagaac 60
|||||
QY 61 acagagacggcggcaggtggtggtccgagtcctggaacgttctctggagagaa 120
|||||
Db 61 acagagacggcggcaggtggtggtccgagtcctggaacgttctctggagagaa 120
|||||
QY 121 caattggcaccatccactatgctgcttggccttggccttggatggacagttggcgtgt 180
|||||
Db 121 caattggcaccatccactatgctgcttggccttggccttggatggacagttggcgtgt 180
|||||
QY 181 atctcagcagtcctcagctcgcctcactagcggggggcgcgcactcctttacgcg 240
|||||
Db 181 atctcagcagtcctcagctcgcctcactagcggggggcgcgcactcctttacgcg 240
|||||
QY 241 atctaatcagtaactcgtctacatctgcatcgtcttctcattagccgaactgaccagc 300
|||||
Db 241 atctaatcagtaactcgtctacatctgcatcgtcttctcattagccgaactgaccagc 300
|||||
QY 301 gttacccagctgcgggtggcacaatactatttggctgcatcgtcagcacaataatc 360
|||||
Db 301 gttacccagctgcgggtggcacaatactatttggctgcatcgtcagcacaataatc 360
|||||
QY 361 aatcgagcattatcagctgctggaactcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 420
|||||
Db 361 aatcgagcattatcagctgctggaactcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 420
|||||
QY 421 agctcagtaactatcactcactcaacagatcccgagcgtcgtcgtcgtcgtcgtcgtcgt 480
|||||
Db 421 agctcagtaactatcactcactcaacagatcccgagcgtcgtcgtcgtcgtcgtcgtcgt 480
|||||

| | | | | |
|--|----|------|--|------|
| | Db | 725 | tgcctcaatcctggcccttcgacataccgtctcaagttcatggcttggaccgcgacgaatg | 784 |
| | Qy | 781 | cattcggctgaagaatgcacagatgctgctcgtacoggtacccaagcagtggttcagtgcac | 840 |
| | Dd | 785 | cattcggctgaagaatgcacagatgctgctcgtacoggtacccaagcagtggttcagtgcac | 844 |
| | Qy | 841 | atcataatcggctctcgaccgccctttccatatacaactcgcagttcttgtatggaaatlaca | 900 |
| | Dd | 845 | atcataatcggctctcgaccgccctttccatatacaactcgcagttcttgtatggaaatlaca | 904 |
| | Qy | 901 | gatctcgactctatctctaagtttcocggcgttatattccattcgcagacaaatgacgcagtc | 960 |
| | Dd | 905 | gatctcgactctatctctaagtttcocggcgttatattccattcgcagacaaatgacgcagtc | 964 |
| | Qy | 961 | cttcggtcgcgtcagttttgcacacggscctctcatatggcggtatctcgatggcctctcttc | 1020 |
| | Dd | 965 | cttcggtcgcgtcagttttgcacacggscctctcatatggcggtatctcgatggcctctcttc | 1024 |
| | Qy | 1021 | gcctccaacgctgtacaagaagactcggtctcgacctcaacctggagctttggcccgggacaat | 1080 |
| | Dd | 1025 | gcctccaacgctgtacaagaagactcggtctcgacctcaacctggagctttggcccgggacaat | 1084 |
| | Qy | 1081 | gggctgggtattttccactcatctcgaaagcattcatccccgctggcaagtctcctggttgg | 1140 |
| | Dd | 1085 | gggctgggtattttccactcatctcgaaagcattcatccccgctggcaagtctcctggttgg | 1144 |
| | Qy | 1141 | tctctattcgcgcacctggggaaattctggccacatgcggatgtatatctcagggttctagc | 1200 |
| | Dd | 1145 | tctctattcgcgcacctggggaaattctggccacatgcggatgtatatctcagggttctagc | 1204 |
| | Qy | 1201 | acagctttaaatgccttggfcaatttcgcgcgttgtaactcagaacactctcctctcagtc | 1260 |
| | Dd | 1205 | acagctttaaatgccttggfcaatttcgcgcgttgtaactcagaacactctcctctcagtc | 1264 |
| | Qy | 1261 | ccaatcgccctactctctaccaaaagcgagatccaaagtctcttgcgcgagcaactcgfgct | 1320 |
| | Dd | 1265 | ccaatcgccctactctctaccaaaagcgagatccaaagtctcttgcgcgagcaactcgfgct | 1324 |
| | Qy | 1321 | tttgtgttacccgcgtggaatcgggtttcttggtcgaattgctagcggtaggtcttcaagtc | 1380 |
| | Dd | 1325 | tttgtgttacccgcgtggaatcgggtttcttggtcgaattgctagcggtaggtcttcaagtc | 1384 |
| | Qy | 1381 | gtcaccaactggttttttcagcttcccaatgacctgacctagcgcgcgcgtccaaccaatgat | 1440 |
| | Dd | 1385 | gtcaccaactggttttttcagcttcccaatgacctgacctagcgcgcgcgtccaaccaatgat | 1444 |
| | Qy | 1441 | tacacaagtgcgattataggcgttgcaacttgcctcttgggtctcttgaaactgggtcggtgat | 1500 |
| | Dd | 1445 | tacacaagtgcgattataggcgttgcaacttgcctcttgggtctcttgaaactgggtcggtgat | 1504 |
| | Qy | 1501 | gccaggaacattatcagggaaccccaacttggagcttgacgcagcgggtcgtcggagcagaa | 1560 |
| | Dd | 1505 | gccaggaacattatcagggaaccccaacttggagcttgacgcagcgggtcgtcggagcagaa | 1564 |
| | Qy | 1561 | tttcaagttcgggccatga | 1578 |
| | Dd | 1565 | tttcaagttcgggccatga | 1582 |

```

RESULT      8
US-09-677-488A-6
: Sequence 6, Application US/09677488A
: GENERAL INFORMATION:
: APPLICANT: Duwick, Jon
: APPLICANT: Maddox, Joyce
: APPLICANT: Gilliam, Jacob
: APPLICANT: Folkerts, Otto
: APPLICANT: Crasta, Oswald R.
: TITLE OF INVENTION: Compositions and Methods
: TITLE OF INVENTION: Detoxification
: FILE REFERENCE: 35718/204100
: CURRENT APPLICATION NUMBER: US/09/677,488A

```

```
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1764
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: permease, partially spliced cDNA
US-09-677-488A-6

Query Match      100.0%; Score 1578; DB 26; Length 1764;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgactccagacacaaatgatacggcgagaaagcgagacagcgagacacgaagac 60
Db 5 atgactccagacacaaatgatacggcgagaaagcgagacagcgagacacgaagac 64
Qy 61 acagagacggcgagcgagtgatcgctccgagtcctcctgaacgttctctggagaagaa 120
Db 65 acagagacggcgagcgagtgatcgctccgagtcctcctgaacgttctctggagaagaa 124
Qy 121 caatttggaccatcacatcgctgcttggccttctgatttgaacagttggcgtgt 180
Db 125 caatttggaccatcacatcgctgcttggccttctgatttgaacagttggcgtgt 184
Qy 181 atctcaggagctccagctgcctactactagcgggggggcccgctactctcttaacgc 240
Db 185 atctcaggagctccagctgcctactactagcgggggggcccgctactctcttaacgc 244
Qy 241 atctcagatcagctctcgtacatctgcctactgccttctcattagcgaactgaccgc 300
Db 245 atctcagatcagctctcgtacatctgcctactgccttctcattagcgaactgaccgc 304
Qy 301 gtctaccgactgcggtggcgaataatcttctgctgatactctcgtgcacacaaatac 360
Db 305 gtctaccgactgcggtggcgaataatcttctgctgatactctcgtgcacacaaatac 364
Qy 361 aatcgagacttcatcagctggtggagctgctgctgttcttcatggatcgactcgga 420
Db 365 aatcgagacttcatcagctggtggagctgctgctgttcttcatggatcgactcgga 424
Qy 421 agctcagtgaccatgatacctgtctcaacagatcccgcgctgatacgccctatagtcac 480
Db 425 agctcagtgaccatgatacctgtctcaacagatcccgcgctgatacgccctatagtcac 484
Qy 481 acatactccagagattcgtggcgtcttctcctcatctacagagagagtcgctggtgtg 540
Db 485 acatactccagagattcgtggcgtcttctcctcatctacagagagagtcgctggtgtg 544
Qy 541 ctcttgttcaactgtttgcccgtgaaagaaaccccttgggttccatgaaatcgattcggc 600
Db 545 ctcttgttcaactgtttgcccgtgaaagaaaccccttgggttccatgaaatcgattcggc 604
Qy 601 ctcagatcgtctctcgtgatactcttctcctcctcctcctcctcctcctcctcctcctc 660
Db 605 ctcagatcgtctctcgtgatactcttctcctcctcctcctcctcctcctcctcctcctc 664
Qy 661 gctccaaactcacaggtatgactgcttggagcaactatactggctggctcgcagcgctc 720
Db 665 gctccaaactcacaggtatgactgcttggagcaactatactggctggctcgcagcgctc 724
Qy 721 tcttctcctggccttgcacatcctgcttctcctcctcctcctcctcctcctcctcctc 780
Db 725 tcttctcctggccttgcacatcctgcttctcctcctcctcctcctcctcctcctcctc 784
Qy 781 catctggctgaagaatgcacagatgctgctcgtctacaggtaccacaaagcagtggtcagtgca 840
```

```
Db 785 catctggctgaagaatgcacagatgctgctcgtacaggtaccacaaagcagtggtcagtgca 844
Qy 841 atcataattggtcttgcaccgcttccatatacacaaatcgcaagttctgtatgaaattaca 900
Db 845 atcataattggtcttgcaccgcttccatatacacaaatcgcaagttctgtatgaaattaca 904
Qy 901 gatctgactctattctaaatcccgcggtctattccattccagacaaatgcagcagtc 960
Db 905 gatctgactctattctaaatcccgcggtctattccattccagacaaatgcagcagtc 964
Qy 961 ctctgctcgtcagtttggcaacggtctctcctcctcctcctcctcctcctcctcctc 1020
Db 965 ctctgctcgtcagtttggcaacggtctctcctcctcctcctcctcctcctcctcctc 1024
Qy 1021 gccctcaacgtctacaagagactgcgtctcctcactcactcagcgttctcccgaggacaat 1080
Db 1025 gccctcaacgtctacaagagagctgcgtctcctcactcactcagcgttctcccgaggacaat 1084
Qy 1081 gggctggtatttccactcactcgcgaacgcatcctcccgctggcgaagtctcctgttgg 1140
Db 1085 gggctggtatttccactcactcgcgaacgcatcctcccgctggcgaagtctcctgttgg 1144
Qy 1141 tctctattcgcaactggtgaattctcgcacatgcgagatgtatatttctagttctcagc 1200
Db 1145 tctctattcgcaactggtgaattctcgcacatgcgagatgtatatttctagttctcagc 1204
Qy 1201 acagcttccaatgcttgggtcaatcccgctgtgtactccagcaactcctcctcctcagtc 1260
Db 1205 acagcttccaatgcttgggtcaatcccgctgtgtactccagcaactcctcctcctcagtc 1264
Qy 1261 ccaatcgccctactcctcactcaaaaagcgagatccaaagtcttctgcgagcactcgtgct 1320
Db 1265 ccaatcgccctactcctcactcaaaaagcgagatccaaagtcttctgcgagcactcgtgct 1324
Qy 1321 ttgtgttaccgctggtgaatcggtttctgttcaatgctgtagcgtgttcttcaagtc 1380
Db 1325 ttgtgttaccgctggtgaatcggtttctgttcaatgctgtagcgtgttcttcaagtc 1384
Qy 1381 gtcacacactggttttccagcttcccaactcagcgtcctcagcccgctcaaccatgaat 1440
Db 1385 gtcacacactggttttccagcttcccaactcagcgtcctcagcccgctcaaccatgaat 1444
Qy 1441 taccaagtcgattatagcgttgcactgctcttgggtgcttgaactgggtcgtgcat 1500
Db 1445 taccaagtcgattatagcgttgcactgctcttgggtgcttgaactgggtcgtgcat 1504
Qy 1501 gccaggaagaattatcagggaccacccacttgagcttgcagcagcgtcgtcagcagca 1560
Db 1505 gccaggaagaattatcagggaccacccacttgagcttgcagcagcgtcgtcagcagca 1564
Qy 1561 ttccaagtggggccatga 1578
Db 1565 ttccaagtggggccatga 1582
```

RESULT 9

```
US-09-677-682A-6
; Sequence 6, Application US/09677682A
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/204101
; CURRENT APPLICATION NUMBER: US/09/677, 682A
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1764
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: permease, partially spliced cDNA
US-09-677-682A-6

Query Match      100.0%; Score 1578; DB 26; Length 1764;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggactccagacaaagtgtatcgacggcgagaaagggcggaagggagagacaaagaaagac 60
Dbb 5 atggactccagacaaagtgtatcgacggcgagaaagggcggaagggagagacaaagaaagac 64
QY 61 acagagacggcgggcgaggtgtgtgcgtccgagtcctcgaacattcctctggaagaaa 120
Dbb 65 acagagacggcgggcgaggtgtgtgcgtccgagtcctcgaacattcctctggaagaaa 124
QY 121 caatttggacattcaacatcgctgtccttggccttggattgtattgcaacagttggcgtggt 180
Dbb 125 caatttggacattcaacatcgctgtccttggccttggattgtattgcaacagttggcgtggt 184
QY 181 atctcaggaagtcctcagctgcgcctactagcggggggggggggggggggggggggggggg 240
Dbb 185 atctcaggaagtcctcagctgcgcctactagcggggggggggggggggggggggggggggg 244
QY 241 atctcaggaagtcctcagctgcgcctactagcggggggggggggggggggggggggggggg 300
Dbb 245 atctcaggaagtcctcagctgcgcctactagcggggggggggggggggggggggggggggg 304
QY 301 gctacccgactgcgggtggcgaataatcttggctgacatcttgcctcgaacaaatacaatc 360
Dbb 305 gctacccgactgcgggtggcgaataatcttggctgacatcttgcctcgaacaaatacaatc 364
QY 361 aatcggagacttcaacgtgtgcgactgcgtgtcgtgttcctcgaacgtgcgtatcgga 420
Dbb 365 aatcggagacttcaacgtgtgcgactgcgtgtcgtgttcctcgaacgtgcgtatcgga 424
QY 421 agctcagtgacatgatacctgtcgaacagatcccgcgctgataccgctcgtatgacac 480
Dbb 425 agctcagtgacatgatacctgtcgaacagatcccgcgctgataccgctcgtatgacac 484
QY 481 acatactccagagattcgtggcattgtcttctcctcattcgaagagagtcgctggtggtg 540
Dbb 485 acatactccagagattcgtggcattgtcttctcctcattcgaagagagtcgctggtggtg 544
QY 541 ctcttggtcaactgtttgcccctgaaagaaaccccttggttcataaactcgagatccggc 600
Dbb 545 ctcttggtcaactgtttgcccctgaaagaaaccccttggttcataaactcgagatccggc 604
QY 601 ctcaacgactcctctcgtgactcctcttctcctcattcgaacattcgaacgctcgaacggc 660
Dbb 605 ctcaacgactcctctcgtgactcctcttctcctcattcgaacattcgaacgctcgaacggc 664
QY 661 gctccaaactcgaagtaggactgcttggaacaaatactggtcgtggttcgaacggcgtc 720
Dbb 665 gctccaaactcgaagtaggactgcttggaacaaatactggtcgtggttcgaacggcgtc 724
QY 721 tgcctcactcgtggccttccgacatcctcgttctcattgcttggttggagcagcaatg 780
Dbb 725 tgcctcactcgtggccttccgacatcctcgttctcattgcttggttggagcagcaatg 784
QY 781 catctggctgaagaaatgcacagatgctgctcgtacaggtaccacaaagcagtggtcagtgca 840
Dbb 785 catctggctgaagaaatgcacagatgctgctcgtacaggtaccacaaagcagtggtcagtgca 844
QY 841 atcataattggctctcgtgacccgcttccatatacaaatcgcgagttctgtatggaaattaca 900

; RESULT 10
; US-09-677-682B-6
; Sequence 6, Application US/09677682B
; GENERAL INFORMATION:
; APPLICANT: DuVick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 35718/204101
; CURRENT FILING DATE: 2000-10-02
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1764
; TYPE: DNA
```

; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: permease, partially spliced cDNA
US-09-677-682B-6

Query Match 100.0%; Score 1578; DB 26; Length 1764;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgagctcagaccagtgatgatacgcgagaaagcgagacagcgagacagacagac 60
Db 5 atgagctcagaccagtgatgatacgcgagaaagcgagacagcgagacagacagac 64
Qy 61 acagagacgagcgagcgagtgatgatacgcgagaaagcgagacagcgagacagac 120
Db 65 acagagacgagcgagcgagtgatgatacgcgagaaagcgagacagcgagacagac 124
Qy 121 caatttggcaccatcacatcagctgctccttggccttggatgttgaacagttggctgggt 180
Db 125 caatttggcaccatcacatcagctgctccttggccttggatgttgaacagttggctgggt 184
Qy 181 atctcagcgagctcagctgctccttgaagcgagggggggcccgctcactctcctttagcgc 240
Db 185 atctcagcgagctcagctgctccttgaagcgagggggggggcccgctcactctcctttagcgc 244
Qy 241 atctcagcgagctcagctgctccttgaagcgagggggggggcccgctcactctcctttagcgc 300
Db 245 atctcagcgagctcagctgctccttgaagcgagggggggggcccgctcactctcctttagcgc 304
Qy 301 gctacccgagctgctcagctgctccttgaagcgagggggggggcccgctcactctcctttagcgc 360
Db 305 gctacccgagctgctcagctgctccttgaagcgagggggggggcccgctcactctcctttagcgc 364
Qy 361 aatcgagcattcagctgctcagctgctccttgaagcgagggggggggcccgctcactctcctttagcgc 420
Db 365 aatcgagcattcagctgctcagctgctccttgaagcgagggggggggcccgctcactctcctttagcgc 424
Qy 421 agctcagtgagcattcagctgctcagctgctccttgaagcgagggggggggcccgctcactctcctttagcgc 480
Db 425 agctcagtgagcattcagctgctcagctgctccttgaagcgagggggggggcccgctcactctcctttagcgc 484
Qy 481 acatactccagagctgctcagctgctccttgaagcgagggggggggcccgctcactctcctttagcgc 540
Db 485 acatactccagagctgctcagctgctccttgaagcgagggggggggcccgctcactctcctttagcgc 544
Qy 541 ctcttgtcagctgctcagctgctccttgaagcgagggggggggcccgctcactctcctttagcgc 600
Db 545 ctcttgtcagctgctcagctgctccttgaagcgagggggggggcccgctcactctcctttagcgc 604
Qy 601 ctcacgagctcctccttgaagcgagggggggggcccgctcactctcctttagcgc 660
Db 605 ctcacgagctcctccttgaagcgagggggggggcccgctcactctcctttagcgc 664
Qy 661 gctcacaactcagagtgagctgctcagctgctccttgaagcgagggggggcccgctcactctcctttagcgc 720
Db 665 gctcacaactcagagtgagctgctcagctgctccttgaagcgagggggggcccgctcactctcctttagcgc 724
Qy 721 tgcctcagctgctcagctgctcagctgctccttgaagcgagggggggcccgctcactctcctttagcgc 780
Db 725 tgcctcagctgctcagctgctcagctgctccttgaagcgagggggggcccgctcactctcctttagcgc 784
Qy 781 catctgctgaagaatgacagagtgctcagctgctccttgaagcgagggggggcccgctcactctcctttagcgc 840
Db 785 catctgctgaagaatgacagagtgctcagctgctccttgaagcgagggggggcccgctcactctcctttagcgc 844
Qy 841 atcataatgctcctcagctgctcagctgctccttgaagcgagggggggcccgctcactctcctttagcgc 900
Db 845 atcataatgctcctcagctgctcagctgctccttgaagcgagggggggcccgctcactctcctttagcgc 904
Qy 901 gatctcagctcctcagctgctcagctgctccttgaagcgagggggggcccgctcactctcctttagcgc 960

Db 905 gatctcagctcctcagctgctcagctgctccttgaagcgagggggcccgctcactctcctttagcgc 964
Qy 961 cttcggtcgctcagctgctcagctgctccttgaagcgagggggcccgctcactctcctttagcgc 1020
Db 965 cttcggtcgctcagctgctcagctgctccttgaagcgagggggcccgctcactctcctttagcgc 1024
Qy 1021 gccctcagctgctcagctgctcagctgctccttgaagcgagggggcccgctcactctcctttagcgc 1080
Db 1025 gccctcagctgctcagctgctcagctgctccttgaagcgagggggcccgctcactctcctttagcgc 1084
Qy 1081 gggctgggtatttccactcctcagctgctcagctgctccttgaagcgagggggcccgctcactctcctttagcgc 1140
Db 1085 gggctgggtatttccactcctcagctgctcagctgctccttgaagcgagggggcccgctcactctcctttagcgc 1144
Qy 1141 tctctattcgcagctgctcagctgctcagctgctccttgaagcgagggggcccgctcactctcctttagcgc 1200
Db 1145 tctctattcgcagctgctcagctgctcagctgctccttgaagcgagggggcccgctcactctcctttagcgc 1204
Qy 1201 acagcttccagctgctcagctgctcagctgctccttgaagcgagggggcccgctcactctcctttagcgc 1260
Db 1205 acagcttccagctgctcagctgctcagctgctccttgaagcgagggggcccgctcactctcctttagcgc 1264
Qy 1261 ccaatcgccctcactcctcagctgctcagctgctccttgaagcgagggggcccgctcactctcctttagcgc 1320
Db 1265 ccaatcgccctcactcctcagctgctcagctgctccttgaagcgagggggcccgctcactctcctttagcgc 1324
Qy 1321 ttgtattcagctgctcagctgctcagctgctccttgaagcgagggggcccgctcactctcctttagcgc 1380
Db 1325 ttgtattcagctgctcagctgctcagctgctccttgaagcgagggggcccgctcactctcctttagcgc 1384
Qy 1381 gtacccactgctgctcagctgctcagctgctccttgaagcgagggggcccgctcactctcctttagcgc 1440
Db 1385 gtacccactgctgctcagctgctcagctgctccttgaagcgagggggcccgctcactctcctttagcgc 1444
Qy 1441 tacacaagtgcgattatagctgctcagctgctccttgaagcgagggggcccgctcactctcctttagcgc 1500
Db 1445 tacacaagtgcgattatagctgctcagctgctccttgaagcgagggggcccgctcactctcctttagcgc 1504
Qy 1501 gccaggaagcattcagctgctcagctgctccttgaagcgagggggcccgctcactctcctttagcgc 1560
Db 1505 gccaggaagcattcagctgctcagctgctccttgaagcgagggggcccgctcactctcctttagcgc 1564
Qy 1561 ttccaagtgggccaatga 1578
Db 1565 ttccaagtgggccaatga 1582

RESULT 11
US-09-882-694-6
; Sequence 6, Application US/09882694
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/208255
; CURRENT APPLICATION NUMBER: US/09/882,694
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1764
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)

OTHER INFORMATION: permease, partially spliced cDNA
US-09-882-694-6

Query Match 100.0%; Score 1578; DB 33; Length 1764;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1578; Conservative 0; Mismatches 0; Indels 0;

[illegible]

| | | | | |
|-----------------------|---|---------------|-----------|--------------|
| Query Match | 100.0% | Score 1578; | DB 33; | Length 1764; |
| Best Local Similarity | 100.0%; | Pred. No. 0; | | |
| Matches 1578; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; |
| Qy 1 | atggactccagaccgaatgggatacggcgagaaagcgggacaagcgagacaacgaagaac | 60 | | |
| Db 5 | atggactccagaccgaatgggatacggcgagaaagcgggacaagcgagacaacgaagaac | 64 | | |
| Qy 61 | acagagacggcgcgacggatgggtgctgcgcaggtccctgaacctctctctggagaagaaa | 120 | | |
| Db 65 | acagagacggcgcgacggatgggtgctgcgcaggtccctgaacctctctctggagaagaaa | 124 | | |
| Qy 121 | caatttggcaccaatcaaccatcgtgtcccttggcctttgtattggcaacagttgggcgtgt | 180 | | |
| Db 125 | caatttggcaaccatcaccatcgtgtcccttggcctttgtattggcaacagttgggcgtgt | 184 | | |
| Qy 181 | atctcaaggcagttcccaagtcgcctactagcaggggggcgccgcaactctctcttacggc | 240 | | |
| Db 185 | atctcaaggcagttcccaagtcgcctactagcaggggggcgccgcaactctctcttacggc | 244 | | |
| Qy 241 | atcctaatacgtactctcgtctacatctgcatacgtcttctcttctcattagccgaactgacgcgc | 300 | | |
| Db 245 | atcctaatacgtactctcgtctacatctgcatacgtcttctcttctcattagccgaactgacgcgc | 304 | | |
| Qy 301 | gtctaccgcactgccgggtggccaatacattttgcgtcgactcctggcaccacaaatcaatc | 360 | | |
| Db 305 | gtctaccgcactgccgggtggccaatacattttgcgtcgactcctggcaccacaaatcaatc | 364 | | |
| Qy 361 | aatcggagcatttctacagttgtgggactcgtgtcgtttgttcttcattggatcgctatcgga | 420 | | |
| Db 365 | aatcggagcatttctacagttgtgggactcgtgtcgtttgttcttcattggatcgctatcgga | 424 | | |
| Qy 421 | agctcagtgacacatgatacctcgtctcaacagatcccgcgctgatagcgcgcctatagtcac | 480 | | |
| Db 425 | agctcagtgacacatgatacctcgtctcaacagatcccgcgctgatagcgcgcctatagtcac | 484 | | |
| Qy 481 | acatactccaggtatcgttgcgactgtcttctctatctacagaggagtcgcgcgtgtgtgtg | 540 | | |
| Db 485 | acatactccaggtatcgttgcgactgtcttctctatctacagaggagtcgcgcgtgtgtgtg | 544 | | |
| Qy 541 | ctcttgttcaacttgtttgcccgtgaaagaaaccccttgggtctatgaaatcgggatctggc | 600 | | |
| Db 545 | ctcttgttcaacttgtttgcccgtgaaagaaaccccttgggtctatgaaatcgggatctggc | 604 | | |
| Qy 601 | ctcacagatcgcctctctctgtgactctcctttatgcgcattctatagcgcgcgtcccaaccacaag | 660 | | |
| Db 605 | ctcacagatcgcctctctctgtgactccttcttctacatctatagcgcgcgtcccaaccacaag | 664 | | |
| Qy 661 | gtctccaaactcacaggtatggactgcttggagcaactatactgctgtgtccgcagcgcctc | 720 | | |
| Db 665 | gtctccaaactcacaggtatggactgcttggagcaactatactgctgtgtccgcagcgcctc | 724 | | |
| Qy 721 | tgttctactcctgggcctttcgacatcctgtctcatgttcttcaattggtctggacgcagcaatg | 780 | | |
| Db 725 | tgttctactcctgggcctttcgacatcctgtctcatgttcttcaattggtctggacgcagcaatg | 784 | | |
| Qy 781 | catctcgtctgaaagatgcacagatgctcgtcgtacggtacccaaagcagtggttcagtgca | 840 | | |
| Db 785 | catctcgtctgaaagatgcacagatgctcgtcgtacggtacccaaagcagtggttcagtgca | 844 | | |
| Qy 841 | atcataattggctttctgcacacgcctttccatatacaatcgcagttctgtatgaaattaca | 900 | | |
| Db 845 | atcataattggctttctgcacacgcctttccatatacaatcgcagttctgtatgaaattaca | 904 | | |
| Qy 901 | gactcgcactctattctaagtctccgcgcgtatattccattctcgagacaatgacgcagctcc | 960 | | |
| Db 905 | gactcgcactctattctaagtctccgcgcgtatattccattctcgagacaatgacgcagctcc | 964 | | |
| Qy 961 | cttcgcgtcgtcgaattttgcaacgctcctcattggtgggtatcgtgatcggcctcttc | 1020 | | |
| Db 965 | cttcgcgtcgtcgaattttgcaacgctcctcattggtgggtatcgtgatcggcctcttc | 1024 | | |
| Qy 1021 | gccctcaacgctgatacaagagactcgtctgcactcactcgtgagctttgcccgggacaat | 1080 | | |

RESULT 13

US-09-351-224-7

03 03 224
: Sequence 7, Application US/09351224

:- GENERAL INFORMATION:

APPLICANT: Duivick Jon

APPLICANT: DUVICK, JOE

TITLE OF INVENTION: Compositions and Methods for Fummonisin Detoxification

FILE OF INVENTION: COMP
FILE REFERENCE: 5718-111

FILE REFERENCE: 5718-111
CURRENT APPLICATION NUMBER: US/09/351 324

1000-07-12

; CURRENT FILING DATE: 1999-07-12

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE:

; SEQ ID NO 7

; LENGTH: 1

TYPE: DNA

ORGANISM:

ORGANISM: *Exophiala spinifrons*
:
: FEATURE:

OTHER INFORMATION:

Query Match

| Query Match | Post local similarity | pred No | score 937; pred No |
|-------------|-----------------------|---------|--------------------|
| 59.46; | 99.78. | | |

Best Local Similarity 99.78; Pred. No. 0;
Matches 1527; Conserving 0; Mismatches 12; Indels 0; Cans 0.

0v 30 aaagccgggacaaagccagacacacagagacggcggcaggtggttc 89

[illegible]

90 cgaatccctgaacattcctctcgaagaagaacaaatttggcaccatccaccatcgtatcctt 149

[illegible]

0y 150 gccctttgtgatttcacacagttagactggtatctcagccagttctccagctgccctact 209

|||||
Db 630 tatcgccatttagcgggtccaaacccaaaggctccaaactcacaggtagtgactgcttg 689
QY gagcaactatactggtctcgacggcgctgctgtcttcacctcggccctttcgacatcctg 749
Db gagcaactatactggtctcgacggcgctgctgtcttcacctcggccctttcgacatcctg 749
QY ctctcatgttcattggtctgacgcagcaaatgcatctggtggaagtgcacagatgctgc 809
Db ctctcatgttcattggtctgacgcagcaaatgcatctggtggaagtgcacagatgctgc 809
QY tcgtacggtagccaaacagtgatgcaatacataaattggtctctgcacgcctttcc 869
Db tcgtacggtagccaaacagtgatgcaatacataaattggtctctgcacgcctttcc 869
QY atatacaatgcagttctgtatggaattacagatctgcactctattctaaagttccgcgg 929
Db atatacaatgcagttctgtatggaattacagatctgcactctattctaaagttccgcgg 929
QY ctatttccattcagagacaatgacgagtcctccttcggtcagttttgcaacggtcct 989
Db ctatttccattcagagacaatgacgagtcctccttcggtcagttttgcaacggtcct 989
QY ctcagtgtggcggtatcgtgatggccttcttcgccctcaacgctgtacaagagactgcgc 1049
Db ctcagtgtggcggtatcgtgatggccttcttcgccctcaacgctgtacaagagactgcgc 1049
QY tcgactcacctggagctttcccgggacaatggcggttatttccactcatctcgaacg 1109
Db tcgactcacctggagctttcccgggacaatggcggttatttccactcatctcgaacg 1109
QY catctaccccgctggcaagtctctgttgggtctctatcgcgacctgggaattctggc 1169
Db catctaccccgctggcaagtctctgttgggtctctatcgcgacctgggaattctggc 1169
QY cacatggcgatgtatatcttaggttcttagcacagcttttcaatgccttggccaattccgc 1229
Db cacatggcgatgtatatcttaggttcttagcacagcttttcaatgccttggccaattccgc 1229
QY cgttgtaactccagcaactctccttcctgatcccaatcgccctactctctacccaaaagcg 1289
Db cgttgtaactccagcaactctccttcctgatcccaatcgccctactctctacccaaaagcg 1289
QY agatccaaaagtcttcggagcactcgtcttctgtttaccgctggaatcggtttct 1349
Db agatccaaaagtcttcggagcactcgtcttctgtttaccgctggaatcggtttct 1349
QY ggtccaatgtgtagcgggtggtcttcacgtccgtcaccaactgtttttcagcttccact 1409
Db ggtccaatgtgtagcgggtggtcttcacgtccgtcaccaactgtttttcagcttccact 1409
QY gacggtgctacgcccgcgtcaacccatgaattacacaagtgcgattataggcgttgcact 1469
Db gacggtgctacgcccgcgtcaacccatgaattacacaagtgcgattataggcgttgcact 1469
QY tgcctcttggtgtcttgaactgggtcgtgcatgccaagagcatttatcagggagcccaactt 1529
Db tgcctcttggtgtcttgaactgggtcgtgcatgccaagagcatttatcagggagcccaactt 1529
QY ggagcttgacggagggctcgtcgagcagaatttcaagttggccatga 1578
Db ggagcttgacggagggctcgtcgagcagaatttcaagttggccatga 1578

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 05:54:29 ; Search time 151.01 Seconds
(without alignments)
2566.781 Million cell updates/sec

Title: US-09-882-694A-7
Perfect score: 1578
Sequence: 1 atgagctccagaccagtg.....aatttcagtggtgcacatga 1578

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2.6/prodata/2/ina/5A_COMB.seq.*
2: /cgn2.6/prodata/2/ina/5B_COMB.seq.*
3: /cgn2.6/prodata/2/ina/6A_COMB.seq.*
4: /cgn2.6/prodata/2/ina/6B_COMB.seq.*
5: /cgn2.6/prodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2.6/prodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------------|
| 1 | 19 | 1.2 | 379 | 1 | US-08-253-155A-51 |
| 2 | 19 | 1.2 | 87350 | 3 | US-08-781-891-79 |
| 3 | 17 | 1.1 | 327 | 1 | US-07-745-382-21 |
| 4 | 17 | 1.1 | 327 | 1 | US-07-921-848-21 |
| 5 | 17 | 1.1 | 327 | 1 | US-08-165-301A-21 |
| 6 | 17 | 1.1 | 327 | 3 | US-08-810-436-21 |
| 7 | 17 | 1.1 | 327 | 5 | PCT-US94-14179-21 |
| 8 | 17 | 1.1 | 330 | 1 | US-08-165-301A-25 |
| 9 | 17 | 1.1 | 330 | 1 | US-08-165-301A-27 |
| 10 | 17 | 1.1 | 330 | 3 | US-08-810-436-25 |
| 11 | 17 | 1.1 | 330 | 3 | US-08-810-436-27 |
| 12 | 17 | 1.1 | 330 | 5 | PCT-US94-14179-25 |
| 13 | 17 | 1.1 | 330 | 5 | PCT-US94-14179-27 |
| 14 | 17 | 1.1 | 1227 | 2 | US-08-683-007A-1 |
| 15 | 17 | 1.1 | 1719 | 4 | US-09-237-111-1 |
| 16 | 17 | 1.1 | 2040 | 3 | US-09-165-042-4 |
| 17 | 17 | 1.1 | 2426 | 4 | US-08-528-026C-3 |
| 18 | 17 | 1.1 | 2472 | 4 | US-09-626-589-6 |
| 19 | 17 | 1.1 | 3348 | 1 | US-08-222-616-35 |
| 20 | 17 | 1.1 | 3348 | 5 | PCT-US95-04228-35 |
| 21 | 17 | 1.1 | 3632 | 1 | US-07-745-382-13 |
| 22 | 17 | 1.1 | 3632 | 1 | US-07-921-848-13 |
| 23 | 17 | 1.1 | 3632 | 1 | US-08-115-680-3 |
| 24 | 17 | 1.1 | 3632 | 1 | US-07-941-372-3 |
| 25 | 17 | 1.1 | 3632 | 1 | US-08-165-301A-13 |
| 26 | 17 | 1.1 | 3632 | 3 | US-08-810-436-13 |
| 27 | 17 | 1.1 | 3632 | 5 | PCT-US93-08247-3 |

| | | | | | | |
|------|----|-----|------|---|-------------------|-------------------|
| C 28 | 17 | 1.1 | 3632 | 5 | PCT-US94-14179-13 | Sequence 13, Appl |
| C 29 | 17 | 1.1 | 4069 | 4 | US-09-302-812-3 | Sequence 3, Appl |
| C 30 | 17 | 1.1 | 4069 | 4 | US-09-511-477-3 | Sequence 3, Appl |
| C 31 | 17 | 1.1 | 4114 | 1 | US-08-165-301A-29 | Sequence 29, Appl |
| C 32 | 17 | 1.1 | 4114 | 3 | US-08-810-436-29 | Sequence 29, Appl |
| C 33 | 17 | 1.1 | 4114 | 5 | PCT-US94-14179-29 | Sequence 29, Appl |
| C 34 | 17 | 1.1 | 4969 | 1 | US-08-260-582-76 | Sequence 76, Appl |
| C 35 | 17 | 1.1 | 4969 | 5 | PCT-US95-05471-76 | Sequence 3, Appl |
| C 36 | 17 | 1.1 | 6619 | 2 | US-08-683-007A-3 | Sequence 33, Appl |
| C 37 | 16 | 1.0 | 273 | 4 | US-08-882-164D-33 | Sequence 110, App |
| C 38 | 16 | 1.0 | 722 | 4 | US-09-328-111-110 | Sequence 656, App |
| C 39 | 16 | 1.0 | 728 | 4 | US-08-998-416-656 | Sequence 1, Appl |
| C 40 | 16 | 1.0 | 836 | 3 | US-08-674-984-2 | Sequence 2, Appl |
| C 41 | 16 | 1.0 | 836 | 3 | US-08-674-984-2 | Sequence 2, Appl |
| C 42 | 16 | 1.0 | 836 | 5 | PCT-US95-15601-1 | Sequence 1, Appl |
| C 43 | 16 | 1.0 | 836 | 5 | PCT-US95-15601-2 | Sequence 22, Appl |
| C 44 | 16 | 1.0 | 1282 | 1 | US-08-361-920-22 | Sequence 22, Appl |
| C 45 | 16 | 1.0 | 1282 | 1 | US-08-479-939-22 | Sequence 22, Appl |

ALIGNMENTS

RESULT 1
US-08-253-155A-51
; Sequence 51, Application US/08253155A
; Patent No. 5691147
; GENERAL INFORMATION:
; APPLICANT: Gyuris, Jeno
; APPLICANT: Draetta, Giulio
; TITLE OF INVENTION: CDR4 Binding Proteins
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/253,155A
; FILING DATE: 02-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: MII-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 379 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-253-155A-51

Query Match 1.2%; Score 19; DB 1; Length 379;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1027 aacgtgtacaagagactg 1045
|||||

DB 100 AACGCTGTACAGAGACTG 118

```

RESULT 2
US-08-781-891-79/c
; Sequence 79, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6090620tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 240052.419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 87350 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-781-891-79

Query Match 1.2%; Score 19; DB 3; Length 87350;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1471 gctcttgatcttgact 1489
Db 57973 GCTCTTGATCTTGAACT 57955

RESULT 3
US-07-745-382-21/c
; Sequence 21, Application US/07745382
; Patent No. 5270181
; GENERAL INFORMATION:
; APPLICANT: McCoy, John
; APPLICANT: Lavallie, Edward
; TITLE OF INVENTION: Peptide and Protein Fusions To
; TITLE OF INVENTION: Thioresoxin and Thioresoxin-Like Molecules
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/921,848
; FILING DATE: 19920728
;
US-07-745-382-21

Query Match 1.1%; Score 17; DB 1; Length 327;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 971 tcagttttgcacggtc 987
Db 178 TCAGTTTGCACGCGTC 162

RESULT 4
US-07-921-848-21/c
; Sequence 21, Application US/07921848
; Patent No. 5292646
; GENERAL INFORMATION:
; APPLICANT: McCoy, John
; APPLICANT: Lavallie, Edward
; TITLE OF INVENTION: Peptide and Protein Fusions To
; TITLE OF INVENTION: Thioresoxin and Thioresoxin-Like Molecules
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/921,848
; FILING DATE: 19920728

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/745,382
; FILING DATE: 19910814
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/652,531
; FILING DATE: 06-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseir, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: G15188A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..327
; OTHER INFORMATION: /citation= ([1])
; PUBLICATION INFORMATION:
; AUTHORS: Lim,
; JOURNAL: J. Bacteriol.
; VOLUME: 163
; PAGES: 311-316
; DATE: 1985
; RELEVANT RESIDUES IN SEQ ID NO: 21: FROM 1 TO 327
;
US-07-745-382-21

Query Match 1.1%; Score 17; DB 1; Length 327;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 971 tcagttttgcacggtc 987
Db 178 TCAGTTTGCACGCGTC 162

RESULT 4
US-07-921-848-21/c
; Sequence 21, Application US/07921848
; Patent No. 5292646
; GENERAL INFORMATION:
; APPLICANT: McCoy, John
; APPLICANT: Lavallie, Edward
; TITLE OF INVENTION: Peptide and Protein Fusions To
; TITLE OF INVENTION: Thioresoxin and Thioresoxin-Like Molecules
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/921,848
; FILING DATE: 19920728

```

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/652,531
; FILING DATE: 06-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/745,382
; FILING DATE: 14-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luano
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: G15188A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..327
; OTHER INFORMATION: /citation= ({1})
; PUBLICATION INFORMATION:
; AUTHORS: Lim,
; JOURNAL: J. Bacteriol.
; VOLUME: 163
; PAGES: 311-316
; DATE: 1985
; RELEVANT RESIDUES IN SEQ ID NO: 21: FROM 1 TO 327
; US-07-921-848-21

Query Match 1.1%; Score 17; DB 1; Length 327;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 971 tcagttttgcaacggtc 987
Db 178 TCAGTTTGGCAACGGTC 162

RESULT 6
US-08-810-436-21/c
; Sequence 21, Application US/08810436
; Patent No. 6143524
; GENERAL INFORMATION:
; APPLICANT: McCoy, John
; APPLICANT: DiBlasio-Smith, Elizabeth
; APPLICANT: Grant, Kathleen
; APPLICANT: LaVallie, Edward R.
; TITLE OF INVENTION: PEPTIDE AND PROTEIN FUSIONS TO
; TITLE OF INVENTION: THIOREDOXIN, THIOREDOXIN-LIKE MOLECULES, AND MODIFIED
; TITLE OF INVENTION: THIOREDOXIN-LIKE MOLECULES
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/810,436
; FILING DATE: 04-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/165,301
; FILING DATE: 10-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meinert, M. C.
; REGISTRATION NUMBER: 33,544
; REFERENCE/DOCKET NUMBER: G1 5188D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/652,531
; FILING DATE: 06-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/745,382
; FILING DATE: 14-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luano
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: G15188A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..327
; OTHER INFORMATION: /citation= ({1})
; PUBLICATION INFORMATION:
; AUTHORS: Lim,
; JOURNAL: J. Bacteriol.
; VOLUME: 163
; PAGES: 311-316
; DATE: 1985
; RELEVANT RESIDUES IN SEQ ID NO: 21: FROM 1 TO 327
; US-07-921-848-21

Query Match 1.1%; Score 17; DB 1; Length 327;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 971 tcagttttgcaacggtc 987
Db 178 TCAGTTTGGCAACGGTC 162

RESULT 5
US-08-165-301A-21/c
; Sequence 21, Application US/08165301A
; Patent No. 5646016
; GENERAL INFORMATION:
; APPLICANT: McCoy, John
; APPLICANT: DiBlasio-Smith, Elizabeth
; APPLICANT: Grant, Kathleen
; APPLICANT: LaVallie, Edward R.
; TITLE OF INVENTION: PEPTIDE AND PROTEIN FUSIONS TO
; TITLE OF INVENTION: THIOREDOXIN, THIOREDOXIN-LIKE MOLECULES, AND MODIFIED
; TITLE OF INVENTION: THIOREDOXIN-LIKE MOLECULES
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/165,301A
; FILING DATE:
; CLASSIFICATION: 435
```

TOPLOGY: unknown
MOLECULE TYPE: DNA
FEATURE: CDS
NAME/KEY: 1.327
LOCATION: 1.327
US-08-810-436-21

Query Match 1.1% Score 17; DB 3; Length 327;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 971 tcagttttgcaacggtc 987
Db 178 TCAGTTTTCACGGTC 162

RESULT 7

PCT-US94-14179-21/c
Sequence 21, Application PC/TUS9414179
GENERAL INFORMATION:
APPLICANT: McCoy, John
APPLICANT: DiBlasio-Smith, Elizabeth
APPLICANT: Grant, Kathleen
APPLICANT: Lavallie, Edward R.
TITLE OF INVENTION: PEPTIDE AND PROTEIN FUSIONS TO
TITLE OF INVENTION: THIOREDOXIN, THIOREDOXIN-LIKE MOLECULES, AND MODIFIED
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/14179
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Meierert, M. C.
REGISTRATION NUMBER: 33,544
REFERENCE/DOCKET NUMBER: GI 5188D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: DNA
MOLECULE TYPE: DNA
FEATURE:

NAME/KEY: CDS
LOCATION: 1.327

PCT-US94-14179-21

Query Match 1.1% Score 17; DB 5; Length 327;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 971 tcagttttgcaacggtc 987
Db 178 TCAGTTTTCACGGTC 162

RESULT 8

US-08-165-301A-25/c
Sequence 25, Application US/08165301A
Patent No. 5646016
GENERAL INFORMATION:
APPLICANT: McCoy, John
APPLICANT: DiBlasio-Smith, Elizabeth
APPLICANT: Grant, Kathleen
APPLICANT: Lavallie, Edward R.
TITLE OF INVENTION: PEPTIDE AND PROTEIN FUSIONS TO
TITLE OF INVENTION: THIOREDOXIN, THIOREDOXIN-LIKE MOLECULES, AND MODIFIED
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/165.301A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meierert, M. C.
REGISTRATION NUMBER: 33,544
REFERENCE/DOCKET NUMBER: GI 5188D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 330 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:

NAME/KEY: CDS
LOCATION: 1.330

US-08-165-301A-25

Query Match 1.1% Score 17; DB 1; Length 330;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 971 tcagttttgcaacggtc 987
Db 178 TCAGTTTTCACGGTC 162

RESULT 9

US-08-165-301A-27/c
Sequence 27, Application US/08165301A
Patent No. 5646016
GENERAL INFORMATION:
APPLICANT: McCoy, John
APPLICANT: DiBlasio-Smith, Elizabeth
APPLICANT: Grant, Kathleen
APPLICANT: Lavallie, Edward R.
TITLE OF INVENTION: PEPTIDE AND PROTEIN FUSIONS TO
TITLE OF INVENTION: THIOREDOXIN, THIOREDOXIN-LIKE MOLECULES, AND MODIFIED
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:


```

; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/165,301A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meinerdt, M. C.
; REGISTRATION NUMBER: 33,544
; REFERENCE/DOCKET NUMBER: GI 5188D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-5851
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 330 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..330
; US-08-165-301A-27

Query Match 1.1% Score 17; DB 1; Length 330;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 971 tcagtttgcaacggtc 987
Db 178 TCAGTTTGCACGGTC 162

RESULT 10
US-08-810-436-25/c
; Sequence 25, Application US/08810436
; Patent No. 6143524
; GENERAL INFORMATION:
; APPLICANT: McCoy, John
; APPLICANT: DiBlasio-Smith, Elizabeth
; APPLICANT: Grant, Kathleen
; APPLICANT: Lavallie, Edward R.
; TITLE OF INVENTION: PEPTIDE AND PROTEIN FUSIONS TO
; TITLE OF INVENTION: THIOREDOXIN, THIOREDOXIN-LIKE MOLECULES, AND MODIFIED
; TITLE OF INVENTION: THIOREDOXIN-LIKE MOLECULES
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/165,301
; FILING DATE: 04-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/165,301
; FILING DATE: 10-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meinerdt, M. C.
; REGISTRATION NUMBER: 33,544
; REFERENCE/DOCKET NUMBER: GI 5188D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-5851
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/165,301
; FILING DATE: 10-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meinerdt, M. C.
; REGISTRATION NUMBER: 33,544
; REFERENCE/DOCKET NUMBER: GI 5188D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 330 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..330
; US-08-810-436-25

Query Match 1.1% Score 17; DB 3; Length 330;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 971 tcagtttgcaacggtc 987
Db 178 TCAGTTTGCACGGTC 162

RESULT 11
US-08-810-436-27/c
; Sequence 27, Application US/08810436
; Patent No. 6143524
; GENERAL INFORMATION:
; APPLICANT: McCoy, John
; APPLICANT: DiBlasio-Smith, Elizabeth
; APPLICANT: Grant, Kathleen
; APPLICANT: Lavallie, Edward R.
; TITLE OF INVENTION: PEPTIDE AND PROTEIN FUSIONS TO
; TITLE OF INVENTION: THIOREDOXIN, THIOREDOXIN-LIKE MOLECULES, AND MODIFIED
; TITLE OF INVENTION: THIOREDOXIN-LIKE MOLECULES
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/810,436
; FILING DATE: 04-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/165,301
; FILING DATE: 10-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meinerdt, M. C.
; REGISTRATION NUMBER: 33,544
; REFERENCE/DOCKET NUMBER: GI 5188D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:

```

LENGTH: 330 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..330
US-08-810-436-27

Query Match 1.1%, Score 17; DB 3; Length 330;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 971 tcagttttgcaacggtc 987
|||||
Db 178 TCAGTTTTCACCGGTC 162

RESULT 12
PCT-US94-14179-25/c
Sequence 25, Application PC/TUS9414179
GENERAL INFORMATION:
APPLICANT: McCoy, John
APPLICANT: DiBlasio-Smith, Elizabeth
APPLICANT: Grant, Kathleen
APPLICANT: Lavallie, Edward R.
TITLE OF INVENTION: PEPTIDE AND PROTEIN FUSIONS TO
THIOREDOXIN, THIOREDOXIN-LIKE MOLECULES, AND MODIFIED
THIOREDOXIN-LIKE MOLECULES
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/14179
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Meinert, M. C.
REGISTRATION NUMBER: 33,544
REFERENCE/DOCKET NUMBER: GI 5188D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 330 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..330
PCT-US94-14179-25

Query Match 1.1%, Score 17; DB 5; Length 330;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 971 tcagttttgcaacggtc 987

Db 178 TCAGTTTTCACCGGTC 162
|||||

RESULT 13
PCT-US94-14179-27/c
Sequence 27, Application PC/TUS9414179
GENERAL INFORMATION:
APPLICANT: McCoy, John
APPLICANT: DiBlasio-Smith, Elizabeth
APPLICANT: Grant, Kathleen
APPLICANT: Lavallie, Edward R.
TITLE OF INVENTION: PEPTIDE AND PROTEIN FUSIONS TO
THIOREDOXIN, THIOREDOXIN-LIKE MOLECULES, AND MODIFIED
THIOREDOXIN-LIKE MOLECULES
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/14179
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Meinert, M. C.
REGISTRATION NUMBER: 33,544
REFERENCE/DOCKET NUMBER: GI 5188D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 330 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..330
PCT-US94-14179-27

Query Match 1.1%, Score 17; DB 5; Length 330;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 971 tcagttttgcaacggtc 987
|||||
Db 178 TCAGTTTTCACCGGTC 162

RESULT 14
US-08-683-007A-1/c
Sequence 1, Application us/08683007A
Patent No. 5858724
GENERAL INFORMATION:
APPLICANT: No. 5858724Y, Robert E
APPLICANT: Domiano, Michael
APPLICANT: Yaeger, Keith
APPLICANT: Kroeker, Warren
TITLE OF INVENTION: Recombinant Rabbit Tissue Factor
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady
 STREET: 1 South Pinckney Street
 CITY: Madison
 STATE: WI
 COUNTRY: USA
 ZIP: 53703
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/683,007A
 FILING DATE: 16-JUL-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Seay, Nicholas J.
 REGISTRATION NUMBER: 27386
 REFERENCE/DOCKET NUMBER: 740380.90040
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 608-251-5000
 TELEFAX: 608-251-9166
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1227 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: Other nucleic acid
 DESCRIPTION: /desc - Recombinant gene encoding
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..1227
 US-08-683-007A-1

Query Match 1.1%; Score 17; DB 2; Length 1227;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 971 tcagtttgcacggtc 987
 Db 178 TCAGTTTTCACACGGTC 162

RESULT 15
 US-09-237-111-1/c
 : Sequence 1, Application US/09237111
 : Patent No. 6340564
 : GENERAL INFORMATION:
 : APPLICANT: Zalacain, Magdalena
 : APPLICANT: Burnham, Martin K. R.
 : APPLICANT: Biswas, Sanjoy
 : APPLICANT: Brown, James R.
 : APPLICANT: Warren, Patrick V.
 : APPLICANT: Ingraham, Karen A.
 : APPLICANT: Chalker, Alison F.
 : APPLICANT: So, Chi-Young
 : APPLICANT: Holmes, David J.
 : APPLICANT: Van Horn, Stephanie
 : APPLICANT: Warren, Richard L.
 : TITLE OF INVENTION: yhx
 : FILE REFERENCE: GM10190
 : CURRENT APPLICATION NUMBER: US/09/237,111
 : CURRENT FILING DATE: 1999-01-26
 : NUMBER OF SEQ ID NOS: 2
 : SOFTWARE: FastSeq for Windows Version 3.0
 : SEQ ID NO 1
 : LENGTH: 1719
 : TYPE: DNA
 : ORGANISM: Streptococcus pneumoniae
 US-09-237-111-1

Query Match 1.1%; Score 17; DB 4; Length 1719;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 427 gtgaccatgatactctgc 443
 Db 425 GTGACCATGATACCTGC 409
 Search completed: April 27, 2002, 05:55:15
 Job time: 15402 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 05:28:23 ; Search time 705 Seconds
(without alignments)
3842.966 Million cell updates/sec

Title: US-09-882-694A-7
Perfect score: 1578
Sequence: 1 atggactccagcaagtgg.....aattccaagttggcccatga 1578

Scoring table: OLIGO.NWC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 937 | 59.4 | 1764 | 22 | AA02693 |
| C 2 | 21 | 1.3 | 11866 | 5 | AA040141 |
| C 3 | 21 | 1.3 | 11873 | 5 | AA040176 |
| C 4 | 21 | 1.3 | 38059 | 22 | AA054018 |
| 5 | 20 | 1.3 | 390 | 21 | AA02802 |
| 6 | 19 | 1.2 | 379 | 17 | AAT12188 |
| 7 | 19 | 1.2 | 718 | 23 | ABL18293 |
| 8 | 19 | 1.2 | 1170 | 22 | AAF63760 |
| C 9 | 19 | 1.2 | 1254 | 17 | AAT31455 |

| | | | | | | |
|------|----|-----|--------|----|-----------|--------------------|
| C 10 | 19 | 1.2 | 2302 | 22 | AAK83590 | Human immune/haema |
| C 11 | 19 | 1.2 | 2718 | 23 | ABL18292 | Drosophila melanog |
| C 12 | 19 | 1.2 | 4519 | 23 | ABL22926 | Drosophila melanog |
| C 13 | 19 | 1.2 | 12332 | 23 | ABL10880 | Drosophila melanog |
| C 14 | 19 | 1.2 | 27869 | 22 | ABLA19635 | Human nervous syst |
| C 15 | 19 | 1.2 | 27869 | 22 | AAK66517 | Human immune/haema |
| C 16 | 19 | 1.2 | 47670 | 23 | ABL16824 | Drosophila melanog |
| C 17 | 19 | 1.2 | 50368 | 23 | ABL16788 | Drosophila melanog |
| C 18 | 19 | 1.2 | 87350 | 18 | AAK83003 | Human WRN genomic |
| C 19 | 18 | 1.1 | 467 | 21 | AAA89575 | Exo19 nucleotide s |
| C 20 | 18 | 1.1 | 555 | 19 | AAV16456 | Endo-beta-1,4-gluc |
| C 21 | 18 | 1.1 | 602 | 22 | AAH87728 | Peppermint plant o |
| C 22 | 18 | 1.1 | 897 | 24 | AB199515 | Mouse ischaemic co |
| C 23 | 18 | 1.1 | 969 | 24 | AA597221 | Neisseria meningit |
| C 24 | 18 | 1.1 | 1219 | 21 | AAA89551 | Mouse syntaxin4 nu |
| C 25 | 18 | 1.1 | 1553 | 23 | AAK86570 | DNA encoding novel |
| C 26 | 18 | 1.1 | 1596 | 17 | AAK22316 | Nocardiopsis sp. P |
| C 27 | 18 | 1.1 | 3000 | 21 | AAA65523 | Porcine BAC-P1Gf2- |
| C 28 | 18 | 1.1 | 3078 | 20 | AAK13360 | Enterococcus faeca |
| C 29 | 18 | 1.1 | 3401 | 23 | ABLO2901 | Drosophila melanog |
| C 30 | 18 | 1.1 | 3447 | 19 | AAV16444 | Glucanase II gene |
| C 31 | 18 | 1.1 | 5421 | 23 | AAK55392 | DNA encoding novel |
| C 32 | 18 | 1.1 | 6007 | 23 | ABLO2900 | Drosophila melanog |
| C 33 | 18 | 1.1 | 6080 | 23 | AAK72880 | DNA encoding novel |
| C 34 | 18 | 1.1 | 168575 | 22 | AAH21613 | Human hyporetin r |
| C 35 | 17 | 1.1 | 193 | 22 | AAK70598 | Human immune/haema |
| C 36 | 17 | 1.1 | 210 | 22 | AAH66973 | C glutamicum codin |
| C 37 | 17 | 1.1 | 229 | 20 | AAV87341 | EST clone BW51. H |
| C 38 | 17 | 1.1 | 249 | 19 | AAV03550 | DNA sequence that |
| C 39 | 17 | 1.1 | 276 | 19 | AAV32676 | Delta thiorodoxin |
| C 40 | 17 | 1.1 | 280 | 19 | AAV32671 | Trxa-concat sequen |
| C 41 | 17 | 1.1 | 282 | 19 | AAV32672 | Delta thiorodoxin |
| C 42 | 17 | 1.1 | 323 | 22 | AAH12667 | Human cDNA clone l |
| C 43 | 17 | 1.1 | 327 | 16 | AAQ90778 | E. coli thiorodoxi |
| C 44 | 17 | 1.1 | 329 | 16 | AAQ90775 | E. coli hprtx gene |
| C 45 | 17 | 1.1 | 329 | 16 | AAQ90776 | E. coli hp2TRX gen |

ALIGNMENTS

RESULT 1
AA02693
ID AAD02693 standard; DNA; 1764 BP.
XX AAD02693;
XX 02-MAY-2001 (first entry)
XX Exophiala spinifera permease DNA.
XX
XX Fumonisin; degradative enzyme; transport enzyme; flavin monooxygenase;
KW aldehyde dehydrogenase; permease; p-glycoprotein; fumonisin esterase;
KW detoxification; mycotoxin; animal feed; human feed; silage;
KW transgenic plant; transgenic animal; microbial spray; ds.
XX Exophiala spinifera.
OS
XX
XX Key Location/Qualifiers
XX CDS 5..1582
XX /tag- a
XX /product- "E. spinifera permease"
XX /note- "This region is specifically claimed in
XX Claim 1b as SEQ ID NO: 7"

WO200105980-A1.
25-JAN-2001.
14-JUL-1999; 99WO-US15824.
14-JUL-1999; 99WO-US15824.


```

RESULT      2
AAN40141/c
ID  AAN40141 standard; DNA; 11866 BP.
XX
XX  AAN40141;
XX
DT  11-FEB-1992 (first entry)
XX
DE  Sequence of human factor IX genomic DNA.
XX
KW  Christmas disease; therapy; haemophilia; factor IX; blood clotting;
XX  diagnosis; ss.
XX
OS  Homo sapiens.
XX
FH  Key      Location/Qualifiers
FT  exon     4441..4560
FT          /*tag= a
FT          /label= AAP40692
FT  exon     7139..7342
FT          /*tag= b
FT          /label= AAP40693
XX
PN  W08400560-A.
XX
PD  16-FEB-1984.
XX
PF  03-AUG-1983; 83WO-GB00191.
XX
PR  06-MAY-1983; 83GB-0012491.
XX  04-AUG-1982; 82GB-0022485.
XX
PA  (NATR ) NATIONAL RES DEV CORP.
XX  (BROW/J) BROWNLEE G G.
XX
PI  Brownlee G, Choo KH;
XX
PI  WPI: 1984-049331/08.
XX  P-PSDB; AAP40692,P40693.
XX
PT  Recombinant DNA cloning vehicles - useful in prodn. of factor IX
XX  polypeptide and of diagnostic probes for Christmas disease
XX
PS  Example: Fig 7a-i; 71pp; English.
XX
CC  The inventors claim DNA molecules comprising part or all of the
XX  human factor IX DNA. The invention also includes cDNA derived from
XX  human factor IX RNA. Specifically claimed are: recombinant DNA (the
XX  phage present in clone lambda HIX-1) deposited as NCIB No. 11749;
XX  Recombinant DNA in which the cloning vehicle is the modified pAT 153
XX  plasmid present in E.coli NCIB No. 11747; Recombinant DNA in which
XX  the bovine factor IX DNA sequence is contained in the recombinant
XX  DNA transformed into E.coli to form a clone deposited as NCIB No.
XX  11748.
XX
SQ  Sequence 11866 BP; 3616 A; 2082 C; 2165 G; 3180 T; 823 other;

Query Match      1.3%; Score 21; DB 5; Length 11866;
Best Local Similarity 100.0%; Pred. NO. 3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  827 cagtgggtcagtgcaatcataa 847
    |||||||
Db  1286 CAGTGGTCAGTGCATCATATAA 1266

RESULT      3
AAN40176/c
ID  AAN40176 standard; DNA; 11873 BP.
XX
XX  AAN40176;
XX
DT  13-FEB-1992 (first entry)
XX
DE  Sequence of human factor IX genomic DNA.
XX
KW  Haemophilia; Christmas disease; diagnosis; treatment; ss.
XX
OS  Homo sapiens.
XX
FH  Key      Location/Qualifiers
FT  CDS      4441..4560
FT          /*tag= a
FT          /note= *see AAP40220*
FT  CDS      7139..7342
FT          /*tag= b
FT          /note= *see AAP40221*
FT  exon     4442..4570
FT          /*tag= c
FT  exon     7140..7342
FT          /*tag= d
FT  repeat_region 7960..8155
FT          /*tag= e
FT  repeat_region 9671..9938
FT          /*tag= f
FT          /rpt_type= Alu1
XX
PN  GB2125409-A.
XX
PD  07-MAR-1984.
XX
PF  03-AUG-1983; 83GB-0020975.
XX
PR  16-MAY-1983; 83GB-0012490.
XX  04-AUG-1982; 82GB-0022486.
XX  03-AUG-1983; 83GB-0020975.
XX
PA  (NATR ) NATIONAL RES DEV CORP.
XX
PI  Brownlee GG, Choo KH;
XX
PI  WPI: 1984-057898/10.
XX  P-PSDB; AAP40220,P40221.
XX
PT  Prodn. of artificial human factor IX - by use of recombinant DNA
XX  sequences for host transformation and cultivation
XX
PS  Example; Fig 7; 49pp; English.
XX
CC  The inventors claim a recombinant DNA having a human factor IX
XX  sequence pref. at least 50 nucleotides long, esp. 75-27000
XX  nucleotides. A cloning vector contg. foreign DNA is also claimed.
XX  The foreign sequence pref. includes the whole of an exon sequence of
XX  the human factor IX genome. The cloning vehicle may be a modified
XX  pAT 153 plasmid. Also claimed is a labelled diagnostic probe
XX  comprising a DNA molecule having a single- or double-stranded probe
XX  sequence of 15 to 10000 nucleotides long Factor IX DNA sequence.
XX
SQ  Sequence 11873 BP; 3615 A; 2078 C; 2170 G; 3180 T; 830 other;

Query Match      1.3%; Score 21; DB 5; Length 11873;
Best Local Similarity 100.0%; Pred. NO. 3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  827 cagtgggtcagtgcaatcataa 847
    |||||||
Db  1286 CAGTGGTCAGTGCATCATATAA 1266

RESULT      4
AAF54018/c
XX
XX  AAF54018/c

```

ID AAF54018 standard; DNA; 38059 BP.
AC AAF54018;
XX 30-MAR-2001 (first entry)
DT Human factor IX (hFIX) gene, SEQ ID NO:4.
DE
XX Age-related gene regulation; liver-specific; gene expression;
KW human factor IX; hFIX; AE5'; AE3'; age-regulatable expression construct;
KW antisense therapy; gene therapy; thrombosis; cardiovascular disease;
KW diabetes; Alzheimer's disease; Parkinson's disease; cancer; osteoporosis;
KW osteoarthritis; dementia; ds.
XX Homo sapiens.
OS
XX WO200075279-A2.
PN
XX 14-DEC-2000.
PD
XX 06-JUN-2000; 2000WO-US15728.
PF
XX 09-JUN-1999; 99US-0328925.
PR
XX (UNMI) UNIV MICHIGAN.
PA Kurachi K, Kurachi S;
PI WPI; 2001-061708/07.
XX P-PSDB; AAB60281, AAB60282, AAB60283, AAB60284, AAB60285, AAB60286,
DR AAB60287, AAB60288, AAB60289.
DR
XX New regulatory elements that control age-related gene expression,
PT useful in gene therapy and for reducing Factor IX expression -
PT
XX Disclosure; Fig 8A-E; 225pp; English.
PS
XX The invention relates to nucleic acid sequences which regulate gene
CC expression in an age-related manner and/or in a liver-specific manner.
CC The invention identifies regions of the human factor IX (hFIX) gene, and
CC a region of the human protein C (hPC) gene, which are age-related
CC regulatory sequences. The hFIX age-related regulatory sequences are
CC designated AE5' (AAF54016) and AE3' (AAF54017) and are found in the 5'
CC UTR (at position 2164-2165 of AAF54018) and 3' UTR (at position
CC 34383-35655 of AAF54018) respectively. These elements act synergistically
CC to increase hFIX levels over the lifespan of an individual; however, they
CC can independently exert effects on hFIX mRNA in an age-related manner,
CC with AE5' acting to stabilise hFIX mRNA, and AE3' acting to increase hFIX
CC mRNA levels, over time. AE5' also directs liver-specific expression. The
CC hPC gene age-related regulatory sequence is found in the 5' UTR
CC (AAF54081) and contains two PEA-3 (polyoma virus activator 3) elements
CC (5'-GAGGAAA-3' and 5'-CAGGAA-3'. The age-related regulatory sequences of
CC the invention, along with their homologues, variants and fragments, may
CC be used in the construction of recombinant expression vectors for the
CC expression of a desired sequence in an age-related fashion in a host
CC cell. Preferred target genes for expression in such age-regulatable
CC expression vectors include those encoding proteins involved in blood
CC coagulation (e.g. the pro-coagulants factor IX and factor VIII, and the
CC anti-coagulants protein C and antithrombin III), human
CC alpha-1-antitrypsin, PEA-3 protein and reporter proteins such as
CC luciferase. Preferred promoters for use in such age-regulatable
CC expression vectors include the human factor IX promoter, the T7 promoter,
CC the T3 promoter and the SP6 promoter. The expression vectors of the
CC invention may be used in gene therapy to provide age-related and/or
CC liver-specific expression of target genes. Age-regulatable constructs may
CC be used in the treatment of such age-related conditions such as
CC thrombosis, cardiovascular disease, diabetes, Alzheimer's disease,
CC Parkinson's disease, cancer, osteoporosis, osteoarthritis and dementia.
CC Specifically, they may be used to express factor IX antisense mRNA in the
CC treatment of thrombotic conditions associated with the natural
CC age-related rise in factor IX expression. Transgenic cells or animals
CC that contain vectors of the invention are useful as models of these
CC diseases, in screening for potential therapeutic agents and for studying

CC normal processes such as ageing and gene expression. Fragments and
CC homologues of age-related regulatory sequences, are useful as probes to
CC detect, isolate or identify other such sequences in samples. The present
CC sequence represents the hFIX gene.
XX
SQ Sequence 38059 BP; 12326 A; 7397 C; 7441 G; 10895 T; 0 other;

Query Match 1.3%; Score 21; DB 22; Length 38059;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 827 cagtgtcagtcgaatcataa 847
|||||
DB 17480 CAGTGTGTCAGTCATCATAA 17460

RESULT 5
AAC02802
ID AAC02802 standard; cDNA; 390 BP.
XX
AC AAC02802;
XX
DT 06-OCT-2000 (first entry)
XX Human secreted protein 5' EST, SEQ ID NO: 2800.
DE
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
KW
XX Homo sapiens.
OS
XX EP1033401-A2.
PN
XX 06-SEP-2000.
PD
XX 21-FEB-2000; 2000EP-0200610.
PF
XX 26-FEB-1999; 99US-0122487.
PR
XX (GEST) GENSET.
PA
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI WPI; 2000-500381/45.
XX P-PSDB; AAG02796.
DR
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
PT
XX Claim 1; SEQ ID 2800; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 390 BP; 78 A; 103 C; 131 G; 76 T; 2 other;

Query Match 1.3%; Score 20; DB 21; Length 390;
Best Local Similarity 100.0%; Pred. No. 11;


```

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 505 gtcttcctcatctacgagg 524
   |||||||
Db 289 gtcttcctcatctacgagg 308

RESULT 6
AAT12188
ID AAT12188 standard; CDNA; 379 BP.
AC AAT12188;
XX
XX 09-AUG-1996 (first entry)
XX
XX pJG4-5-CDK-BP cDNA clone #68 3' fragment.
XX
XX Cell cycle; regulation; G1 phase; proliferation; tumorigenesis;
XX cyclin dependent kinase; differentiation; CDK4 inhibitor; agonist;
XX antagonist; ss.
XX
XX Synthetic.
XX
XX WO9533819-A2.
XX
XX 14-DEC-1995.
XX
XX 02-JUN-1995; 95WO-US07113.
XX
XX 02-JUN-1994; 94US-0253155.
XX
XX (MITO-) MITOTIX INC.
XX
XX Draetta G, Gyuris J;
XX
XX WPI; 1996-040227/04.
XX
XX Cyclin-dependent kinase-4 binding protein - used in the isolation of
XX (ant)agonists of cell cycle regulation.
XX
XX Claim 30; Page 93; 115pp; English.
XX
XX AAT12186-T12207 are internal or 3' fragments of cDNA clones of the
XX plasmid pJG4-5-CDK-BP. The 5' end of the clones encode cyclin dependent
XX kinase 4 (CDK4) binding proteins (CDK4-BP), which may be used in an assay
XX for screening test compounds as inhibitors of CDK/CDK4-BP interaction.
XX The complexes formed by CDK4 and D-type cyclins are strongly implicated
XX in the control of the early G1 phase of the cell cycle and are strong
XX candidates for controlling and/or preventing tumorigenesis and the
XX onset of cancer. Nucleic acids encoding CDK4-BP or fragments of these
XX may be used as probes/primers to diagnose the presence or absence of
XX genetic lesions in a gene encoding 1 of 24 claimed CDK4-BP, and hence to
XX diagnose the risk for a subject of developing a cell-proliferation
XX associated disorder (e.g. cancer).
XX
XX Sequence 379 BP; 105 A; 96 C; 80 G; 96 T; 2 other;

Query Match 1.2%; Score 19; DB 17; Length 379;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1027 aacgctgtacaagagactg 1045
   |||||||
Db 100 aacgctgtacaagagactg 118

RESULT 7
ABL18293
ID ABL18293 standard; DNA; 718 BP.
XX
XX ABL18293;
XX

```

```

DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 6352.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 6352; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 718 BP; 176 A; 172 C; 164 G; 206 T; 0 other;

Query Match 1.2%; Score 19; DB 23; Length 718;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 890 atggaattacagatctcga 908
   |||||||
Db 218 atggaattacagatctcga 236

RESULT 8
AAF63760
ID AAF63760 standard; DNA; 1170 BP.
XX
XX AAF63760;
XX
XX 03-APR-2001 (first entry)
XX
XX Drosophila gustatory receptor GR59D.1 DNA sequence.
XX
XX Gustatory receptor; fruit fly; taste; pheromone; semiochemical;
XX crop damage; pest control; ds.
XX
XX Drosophila melanogaster.
XX
XX WO200077208-A2.
XX
XX 21-DEC-2000.
XX

```

PF 14-JUN-2000; 2000WO-US16211.
 XX
 PR 14-JUN-1999; 99US-0138668.
 PR 10-FEB-2000; 2000US-0181704.
 XX
 XX (UYVA) UNIV YALE.
 PA
 PI Carlsson PJ, Clyne PJ, Warr CG;
 XX P-PSDB; AAB75221.
 DR WPI: 2001-061873/07.
 DR
 XX
 PT New isolated nucleic acid molecule encoding Drosophila Gustatory
 PT Receptor protein useful for e.g. identification of compounds which may
 PT be used for pest management -
 XX
 XX Claim 3; Page 172-174; 227pp; English.
 XX
 CC This invention relates to polynucleotide sequences AAF63732 - AAF63777
 CC which encode Drosophila gustatory receptor proteins represented by
 CC sequences AAB75193 - AAB75238. The invention includes methods for
 CC determining gustatory receptor ligands. Also included is a method for
 CC modulating the expression of the DNA encoding the receptors. The DNA and
 CC protein sequences may be used for the identification of compounds,
 CC e.g. pheromones and other semiochemicals, which may be used for pest
 CC management. The DNA sequences may also be used for behavioural studies
 CC involving gustatory systems in various organisms. Also, the DNA sequences
 CC may also be used to track down gustatory receptor genes in insects that
 CC damage crops or transmit diseases.
 XX
 SQ Sequence 1170 BP; 275 A; 275 C; 285 G; 335 T; 0 other;

Query Match 1.2%; Score 19; DB 22; Length 1170;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1267 gccctactctctaccacaa 1285
 |||||
 DB 154 gccctactctctaccacaa 172

RESULT 9
 AAT31455/c
 ID AAT31455 standard; DNA; 1254 BP.

XX
 AC AAT31455;
 XX
 DT 15-SEP-1996 (first entry)
 XX
 DE CDK6 inhibiting protein gene.
 XX
 KW CDK6 inhibiting protein; cyclin-dependent kinase; tumour;
 KW gene therapy; antisense; p20; ss.
 XX
 OS Homo sapiens.

FH Key Location/Qualifiers
 FT CDS 140..640
 /*tag= a

XX WO9619244-A1.
 XX
 PD 27-JUN-1996.
 XX
 PF 19-DEC-1995; 95WO-US16553.
 XX
 PR 22-DEC-1994; 94US-0362235.
 XX
 PA (UNMI) UNIV MICHIGAN.
 PA (UYNC-) UNIV NORTH CAROLINA.
 XX
 PI Guan K, Xiong Y;

XX WPI: 1996-309319/31.
 DR P-PSDB; AAR98367.
 XX
 PT Nucleic acid encoding cyclin-dependent kinase-inhibiting protein -
 PT for inhibiting DNA synthesis in tumour cells in vitro, and for
 PT designing antisense oligonucleotide(s) for increasing DNA
 PT synthesis in a cell
 XX
 XX Claim 1; Page 27-28; 37pp; English.
 XX
 CC A DNA sequence (AAT31455) codes for p20 (AAR98367), a 20 kDa inhibitor
 CC of cyclin-dependent kinase 6 (CDK6). It was isolated from a
 CC human HeLa cDNA library by screening with a fragment amplified by
 CC PCR from cDNA using degenerate primers based on the conserved
 CC regions of p16 and p18 (see also AAR98368) protein sequences.
 CC The gene can be used to design hybridisation probes, or in gene
 CC therapy methods for the inhibition of DNA synthesis in tumour cells.
 CC Antisense sequences can be used to increase DNA synthesis in skin
 CC cells in wound or burn tissue.
 XX
 SQ Sequence 1254 BP; 236 A; 326 C; 396 G; 296 T; 0 other;

Query Match 1.2%; Score 19; DB 17; Length 1254;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1027 aacgctgtacaagagactg 1045
 |||||
 DB 1142 AACGCTGTACAAGAGACTG 1124

RESULT 10
 AAK83590/c
 ID AAK83590 standard; DNA; 2302 BP.

XX
 AC AAK83590;
 XX
 DT 07-NOV-2001 (first entry)
 XX
 DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:38402.
 XX
 KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytostatic; gene therapy; vaccine; metastasis; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200157182-A2.
 XX
 PD 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US01354.
 XX
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226691.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 23-AUG-2000; 2000US-0228924.
PR 30-AUG-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis -

Disclosure; SEQ ID NO 38402; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patients own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I), by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, diagnose and treat immune/haematopoietic-related diseases, especially cancers and cancer metastases of haematopoietic-derived cells. AAK64703 to AAK87694 represent human immune/haematopoietic antigen genomic sequences from the present invention. AAK54942 to AAK54950 and AAK82169

CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 2302 BP; 401 A; 572 C; 712 G; 617 T; 0 other;

Query Match 1.2%; Score 19; DB 22; Length 2302;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1027 aacgctgtacaagagactg 1045
|||||
Db 2204 AACGCTGTACAGAGACTG 2186

RESULT 11
ABL18292/c
ID ABL18292 standard; DNA; 2718 BP.

XX ABL18292;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 6349.

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions -

XX Claim 1; SEQ ID NO 6349; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
sequences (ABL01840-ABL16175) and the encoded proteins
(ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2718 BP; 803 A; 546 C; 582 G; 787 T; 0 other;

Query Match 1.2%; Score 19; DB 23; Length 2718;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 890 atggaattacagatctcga 908
|||||

Db 1501 ATGGAATTACAGATCTCGA 1483

RESULT 12

ABL22926/c
ID ABL22926 standard; DNA; 4519 BP.

XX ABL22926;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 20251.

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions -

XX Claim 1; SEQ ID NO 20251; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
sequences (ABL01840-ABL16175) and the encoded proteins
(ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 4519 BP; 1283 A; 1005 C; 1023 G; 1208 T; 0 other;

Query Match 1.2%; Score 19; DB 23; Length 4519;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1267 gccctactctctaccacaa 1285
|||||

Db 3540 GCCCTACTCTCTACCAA 3522

RESULT 13

ABL10880
ID ABL10880 standard; cDNA; 12332 BP.

XX ABL10880;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 27122.

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.

```
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-101637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR P-PSDB; ABB66777.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Claim 1; SEQ ID NO 27122; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 12332 BP; 3452 A; 2580 C; 2685 G; 3615 T; 0 other;

Query Match 1.2%; Score 19; DB 23; Length 12332;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 260 tctacatctgcgcgttt 278
|||||
Db 6219 tctacatctgcgcgttt 6237

RESULT 14
ABAI9635
ID ABAI9635 standard; DNA; 27869 BP.
XX AC ABAI9635;
XX DT 23-JAN-2002 (first entry)
XX DE Human nervous system related polynucleotide SEQ ID NO 11966.
XX KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
XX KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
XX KW antiparkinsonian; antiscikling; antianaemic; antiarthritic; cancer;
XX KW antifuramatic; hepatotropic; cerebroprotective; antiinflammatory;
XX KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
XX KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
XX KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX OS Homo sapiens.
XX PN WO200159063-A2.
XX PD 16-AUG-2001.
XX PR 17-JAN-2001; 2001WO-US01334.
XX PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.
XX PR 19-MAY-2000; 2000US-0205515.
XX PR 07-JUN-2000; 2000US-0209467.
XX PR 28-JUN-2000; 2000US-0214886.
XX PR 30-JUN-2000; 2000US-0215135.
XX PR 07-JUL-2000; 2000US-0216647.
XX PR 11-JUL-2000; 2000US-0216880.
XX PR 11-JUL-2000; 2000US-0217487.
XX PR 11-JUL-2000; 2000US-0217496.
XX PR 14-JUL-2000; 2000US-0218290.
XX PR 26-JUL-2000; 2000US-0220963.
XX PR 26-JUL-2000; 2000US-0220964.
XX PR 14-AUG-2000; 2000US-0224518.
XX PR 14-AUG-2000; 2000US-0224519.
XX PR 14-AUG-2000; 2000US-0225213.
XX PR 14-AUG-2000; 2000US-0225214.
XX PR 14-AUG-2000; 2000US-0225266.
XX PR 14-AUG-2000; 2000US-0225267.
XX PR 14-AUG-2000; 2000US-0225268.
XX PR 14-AUG-2000; 2000US-0225270.
XX PR 14-AUG-2000; 2000US-0225447.
XX PR 14-AUG-2000; 2000US-0225757.
XX PR 14-AUG-2000; 2000US-0225758.
XX PR 14-AUG-2000; 2000US-0225759.
XX PR 18-AUG-2000; 2000US-0226279.
XX PR 22-AUG-2000; 2000US-0226681.
XX PR 22-AUG-2000; 2000US-0226868.
XX PR 23-AUG-2000; 2000US-0227182.
XX PR 30-AUG-2000; 2000US-0227009.
XX PR 01-SEP-2000; 2000US-0228924.
XX PR 01-SEP-2000; 2000US-0229387.
XX PR 01-SEP-2000; 2000US-0229343.
XX PR 01-SEP-2000; 2000US-0229344.
XX PR 05-SEP-2000; 2000US-0229345.
XX PR 05-SEP-2000; 2000US-0229509.
XX PR 06-SEP-2000; 2000US-0229513.
XX PR 06-SEP-2000; 2000US-0230437.
XX PR 08-SEP-2000; 2000US-0230438.
XX PR 08-SEP-2000; 2000US-0231242.
XX PR 08-SEP-2000; 2000US-0231243.
XX PR 08-SEP-2000; 2000US-0231244.
XX PR 08-SEP-2000; 2000US-0231413.
XX PR 08-SEP-2000; 2000US-0231414.
XX PR 08-SEP-2000; 2000US-0232080.
XX PR 12-SEP-2000; 2000US-0232081.
XX PR 14-SEP-2000; 2000US-0232397.
XX PR 14-SEP-2000; 2000US-0232398.
XX PR 14-SEP-2000; 2000US-0232399.
XX PR 14-SEP-2000; 2000US-0232400.
XX PR 14-SEP-2000; 2000US-0232401.
XX PR 14-SEP-2000; 2000US-0233063.
XX PR 14-SEP-2000; 2000US-0233064.
XX PR 14-SEP-2000; 2000US-0233065.
XX PR 21-SEP-2000; 2000US-0234223.
XX PR 21-SEP-2000; 2000US-0234274.
XX PR 25-SEP-2000; 2000US-0234997.
XX PR 26-SEP-2000; 2000US-0234998.
XX PR 27-SEP-2000; 2000US-0235484.
XX PR 27-SEP-2000; 2000US-0235834.
XX PR 27-SEP-2000; 2000US-0235836.
XX PR 29-SEP-2000; 2000US-0236327.
XX PR 29-SEP-2000; 2000US-0236367.
XX PR 29-SEP-2000; 2000US-0236368.
XX PR 29-SEP-2000; 2000US-0236369.
```

29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0242221.
 PR 08-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0246617.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 XX
 PA Rosen CA, Barash SC, Ruben SM;
 PI WPI; 2001-541565/60.
 XX
 DR Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 XX
 PT

useful for preventing, diagnosing and/or treating nervous system
 cancers and metastases -
 PT
 XX
 XX Disclosure; SEQ ID NO 11966; 1701pp + Sequence listing; English.
 XX
 CC The invention relates to novel genes (ABAI1004-ABA21534) and proteins
 (ABH14678-ABH18001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 27869 BP; 7054 A; 6673 C; 6322 G; 7820 T; 0 other;

Query Match 1.2%; Score 19; DB 22; Length 27869;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 ttgtgattgcacagttg 173
 |||||
 Db 26107 ttgtgattgcacagttg 26125

RESULT 15
 AAK66517/c
 ID AAK66517 standard; DNA: 27869 BP.
 XX
 AC AAK66517;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:21329.
 XX
 KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytostatic; gene therapy; vaccine; metastasis; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200157182-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01354.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.

CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 27869 BP; 7820 A; 6322 C; 6673 G; 7054 T; 0 other;

Query Match 1.2%; Score 19; DB 22; Length 27869;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 155 ttgtgatttgcaacagttg 173
|||||
Db 1763 TTGTGATTTCACAGTTG 1745

Search completed: April 27, 2002, 05:29:34
Job time: 13857 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 05:34:37 ; Search time 7820.58 Seconds
(without alignments)
4222.459 Million cell updates/sec

Title: US-09-882-694A-7

Perfect score: 1578

Sequence: 1 atggactccagaccagtg.....aattcaagtgggscatga 1578

Scoring table: OLIGO.NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pi.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Match | Length | ID | Description |
|------------|-------------|--------|----|-------------|
|------------|-------------|--------|----|-------------|

| | | | | | | |
|----|-----|------|--------|----|----------|---------------------|
| 1 | 937 | 59.4 | 1578 | 6 | AX076848 | AX076848 Sequence |
| 2 | 937 | 59.4 | 1764 | 6 | AX076847 | AX076847 Sequence |
| 3 | 22 | 1.4 | 38209 | 9 | AF000292 | AF000292 Homo sapi |
| 4 | 22 | 1.4 | 40697 | 9 | AF039905 | AF039905 Homo sapi |
| 5 | 22 | 1.4 | 100000 | 9 | AF000043 | AF000043 Homo sapi |
| 6 | 22 | 1.4 | 100000 | 9 | AF000111 | AF000111 Homo sapi |
| 7 | 22 | 1.4 | 100000 | 9 | AF000187 | AF000187 Homo sapi |
| 8 | 22 | 1.4 | 340000 | 9 | AF001716 | AF001716 Homo sapi |
| 9 | 21 | 1.3 | 11873 | 6 | HSFAC1XG | A2491 H.sapiens f |
| 10 | 21 | 1.3 | 38059 | 9 | HUMFIXG | AL02402 Human coagu |
| 11 | 21 | 1.3 | 153615 | 2 | AL645665 | AL045665 Homo sapi |
| 12 | 21 | 1.3 | 158557 | 9 | HS88D7 | AL033403 Human DNA |
| 13 | 20 | 1.3 | 2154 | 8 | AF052582 | AF052582 Fusarium |
| 14 | 20 | 1.3 | 2770 | 10 | RNPFKL | X58865 Rat PFK-L m |
| 15 | 20 | 1.3 | 2879 | 9 | BC007536 | BC007536 Homo sapi |
| 16 | 20 | 1.3 | 2914 | 9 | HSPFKLA | X15573 Human liver |
| 17 | 20 | 1.3 | 2919 | 9 | BC008964 | BC008964 Homo sapi |
| 18 | 20 | 1.3 | 2920 | 9 | BC009919 | BC009919 Homo sapi |
| 19 | 20 | 1.3 | 3385 | 9 | BC006422 | BC006422 Homo sapi |
| 20 | 20 | 1.3 | 14023 | 1 | AE007062 | AE007062 Mycobacte |
| 21 | 20 | 1.3 | 46778 | 9 | AL356003 | AL356003 Human DNA |
| 22 | 20 | 1.3 | 53450 | 1 | MTV018 | AL021899 Mycobacte |
| 23 | 20 | 1.3 | 81908 | 9 | AL590623 | AL590623 Human DNA |
| 24 | 20 | 1.3 | 83408 | 8 | AE078516 | AB078516 Arabidops |
| 25 | 20 | 1.3 | 85382 | 9 | AL390793 | AL390793 Human DNA |
| 26 | 20 | 1.3 | 115890 | 2 | AP02824 | AP02824 Homo sapi |
| 27 | 20 | 1.3 | 153094 | 2 | AC091973 | AC091973 Homo sapi |
| 28 | 20 | 1.3 | 153819 | 2 | AC109466 | AC109466 Homo sapi |
| 29 | 20 | 1.3 | 157078 | 9 | AC096727 | AC096727 Homo sapi |
| 30 | 20 | 1.3 | 179415 | 2 | AC024286 | AC024286 Homo sapi |
| 31 | 20 | 1.3 | 18872 | 9 | AC020760 | AC020760 Homo sapi |
| 32 | 20 | 1.3 | 193910 | 2 | AC040956 | AC040956 Homo sapi |
| 33 | 20 | 1.3 | 232506 | 2 | AC109822 | AC109822 Homo sapi |
| 34 | 19 | 1.2 | 177 | 9 | HSPFK02 | X16912 Human PFKL |
| 35 | 19 | 1.2 | 379 | 6 | I76223 | I76223 Sequence 51 |
| 36 | 19 | 1.2 | 955 | 5 | TCU94815 | U94815 Trogon curu |
| 37 | 19 | 1.2 | 955 | 5 | TVU94816 | U94816 Trogon viol |
| 38 | 19 | 1.2 | 1155 | 9 | BC001822 | BC001822 Homo sapi |
| 39 | 19 | 1.2 | 1170 | 6 | AX058757 | AX058757 Sequence |
| 40 | 19 | 1.2 | 1256 | 9 | HSU40343 | U40343 Human CDK i |
| 41 | 19 | 1.2 | 1901 | 3 | AY071767 | AY071767 Drosophi |
| 42 | 19 | 1.2 | 3470 | 3 | AY052145 | AY052145 Drosophi |
| 43 | 19 | 1.2 | 6881 | 1 | PFUTRAID | M54796 Plasmid F (|
| 44 | 19 | 1.2 | 15847 | 9 | AF001061 | AF001061 Homo sapi |
| 45 | 19 | 1.2 | 16928 | 2 | AC109117 | AC109117 Rattus no |

ALIGNMENTS

| | | | | | | |
|------------|--|--------------------------|-------|--------|-----------------|--|
| RESULT | 1 | | | | | |
| AX076848 | | | | | | |
| LOCUS | AX076848 | 1578 bp | DNA | linear | PAT 22-FEB-2001 | |
| DEFINITION | Sequence | 7 from Patent WO0105980. | | | | |
| ACCESSION | AX076848 | | | | | |
| VERSION | AX076848.1 | GI:13121521 | | | | |
| KEYWORDS | | | | | | |
| SOURCE | | | | | | |
| ORGANISM | | | | | | |
| | Exophiala spinifera. | | | | | |
| | Exophiala spinifera | | | | | |
| | Eukaryota; Fungi; Ascomycota; Pezizomycotina; Chaetothyriomycetes; | | | | | |
| | Chaetothyriales; Herpotrichiellaceae; mitosporic | | | | | |
| | Herpotrichiellaceae; Exophiala. | | | | | |
| REFERENCE | 1 (bases 1 to 1578) | | | | | |
| AUTHORS | Duvick J.P., Maddox J., Gilliam J., Folkerts O. and Crasta, O.R. | | | | | |
| TITLE | Compositions and methods for fumonisin detoxification | | | | | |
| JOURNAL | Patent: WO 0105980-A 7 25-JAN-2001; | | | | | |
| FEATURES | Pioneer Hi-Bred International Inc. (US) ; Curagen Corporation (US) | | | | | |
| | Location/Qualifiers | | | | | |
| | source | | | | | |
| | i. .1578 | | | | | |
| | /organism="Exophiala spinifera" | | | | | |
| | /db_xref="taxon:91928" | | | | | |
| | /note="permease, coding sequence" | | | | | |
| BASE COUNT | 324 a | 446 c | 385 g | 423 t | | |

Q:

Query Match 59.4%; Score 937; DB 6; Length 1764;
Best Local Similarity 99.2%; Pred. No. 0;

Best Local Similarity 99.2%; Pred. NO: 0;
Matches 1537; Conservative 12; Indels 0; Gaps 0;

RESULT

DNA linear PAT 22-FEB-2001

an eni nifera

a spinifera

aceae; mitosporic

is 1 to 1764;

J.J., Folkerts, O. and Craske, O.R.:
monisin detoxification

1001: 1001.

Location/Qu
1. .1764

```
/organism="
/db xref="t
```

/note="perm
a 478 c

EO 48

ity 99.28;

PARTIALS

acvayucavayucav

ACAAGGCAGACA

QY 90 cgaagtcctgaacgttctctctgagaagaacaattggcaccatcaccatcggtctctt 149
DB 94 CGAGTCCCTGAACGTTCCCTCTGGAGAAGAAACAATTTGGACCAATCAACATCGTGTCTT 153
QY 150 ggccttgatgattgaacagttggctggtatctcagcagctctccagctcgccctact 209
DB 154 GGCCTTGTGATTTGCAACAGTTGGCTGGTATCTCAGGCACTCTCCAGCTCGCCCTACT 213
QY 210 agcggggggcccgctcaactctctcttaacggcatccttaataatcagctactctctcaactg 269
DB 214 AGCGGGGGGGCCCGCTCACTCTCTCTTACGGCATCTTAATCAGTACTCTGCTCTACATCTG 273
QY 270 catcgcttctctattagcgaactaacagcgtctaccgcagctcggtgcccgaatataca 329
DB 274 CATCGCTTTCTCATTTAGCGAACTGACCAAGCTCTACCGAGCTCGCGGTGGCAATATCA 333
QY 330 ttttggtgatctctggcaccacaaatcaatcaatcgagcattctcaatcggtgagact 389
DB 334 TTTTGGCTCGATCTGGCACCACAAATCAATCAATCGGAGCATTTTCATACGTGTGGGACT 393
QY 390 cgtgctgcttctctaatcgatcgctatcggaagctcaagctcaatgacatgactcgctcaaca 449
DB 394 CGTGTGCTTCTTTCATGATCTCTATCGGAAGCTCAGTGACCATGATACCTGCTCAACA 453
QY 450 gatccggcgctgatagcgcgcttatagtcacacatactctccaggattcggtgcatgctt 509
DB 454 GATCCCGGGCTGATAGCGCCCTATAGTCACACATACTCCAGGATTCGTGGCATGTCTT 513
QY 510 cctcatcagagggagtcgctggtgctctgttctgttcaactgtttgcccgaagaag 569
DB 514 CCTCATCTCAGGGAGTCGCGTGGTGTGCTTGTGTTCAACTTGTGTGCCCTGAAGA 573
QY 570 aacccctgggttcatagaatcgattcgctcagatcgctctctctctgctgactcctt 629
DB 574 AAACCCCTTGCTCATGAATCGAATTCGGCCTTCACGATCGCTCTCTCTCGTATCTCCTT 633
QY 630 tatgcacatctagcgcgtccaacccaaagctccaaactcaaggtatgagactgcttg 689
DB 634 TATCGCCATTTCTAGCGGGTCCAAACCCCAAGGCTCCAAACTCACAGGTATGGACTGCTTG 693
QY 690 gagcaactactggctggtccgaacgctgctgtcttctcctctggccttctgacatcctg 749
DB 694 GAGCAACTATATCTGGTGTCCGACGCGCTCTCTCTCATCTCTGGGCTTTGACATCTCTG 753
QY 750 ctctcatctctgcttgagcagcaaatgactcgtcggtgaagaatgacagatgctgc 809
DB 754 CTTTCATGTTCTATGCTGTGGACGACGAATGCAATCTGCTGGCTGAAGATGACAGATGCTGC 813
QY 810 tctgacggtaccacaaagcagtggtcagtgcaataatgtgctctgacccgctcttc 869
DB 814 TCGTACGTTACCCAAAGCACTGCTCAGTGCATCAATCAATTTGGCTTCTGCACCGCTTTCC 873
QY 870 atatacaatcgagttctgtatgaattacagatctcgactctattcttaagttccgcgg 929
DB 874 ATATACAATCGAGTCTGTATGGAATACAGATCTCGACTCTATTCTAAGTTCCGCCGG 933
QY 930 ctatctccatctgagcaaatgacgcatctcttcctggctgctcagttttgcaacggtcct 989
DB 934 CTATATCTCATCTCGAGCAATAGGCGAGTCTGCTCGAATTCGGCTTTTTCGAACGCTCT 993
QY 990 ctcagtggcggtatctgtagtccttctctgcccctcaacgctgtacaaagactgctc 1049
DB 994 CTCATGTGGCGGTATCTGTATGCTCTTCTTCGCCCTCAACGCTGTACAAAGAGACTGCTC 1053
QY 1050 tccgactcacctggagcttggccgggacaatggcgtgggtattttccactcatctcgaaag 1109
DB 1054 TCGACTCACCTGGAGCTTTGCCGGGACATGGCTGGTATTTTCCACTCATCTCGAAGC 1113
QY 1110 catctcctcccgctggaagttctctgtttggtctctattcgagacctgggggaattctggc 1169
DB 1114 CATTCATCCCGCTGGCAAGTTCTCTGTTGGTCTCTATTTCGGACCTGGGGAATTCGTGC 1173
QY 1170 cacatggcgatgatattcttagttcttagcacagacttcaatgccttggctgaattccgc 1229

DB 1174 CACATCGGATGTATATTTCTAGGTCTAGCACAGCTTTCAATGCTTGGTCAATTCGCG 1233
QY 1230 cgttgactccagcaactctctctctgactcccaatcgccctactctctaccacaaagcg 1289
DB 1234 CGTTGTACTCCAGCAACTCTCTCTCTGATCCCAATCGCCCTACTCTCTACCAAAAGCG 1293
QY 1290 agatccaaagtcttccgagcaactcgtgctttttgtttgttacgcggtggaatcggtttct 1349
DB 1294 AGATCCAAAGTTCTTCCGAGCACTCGTCTTTTGTGTTTACCGGTGGAATCGGGTTTCT 1353
QY 1350 ggtcaatgctgactagcgggtgtcttccgctccgctcaccactgtgtttttcaagcttccca 1409
DB 1354 GGTCAATGTGCTAGCGGTGTCTTTCAGTCCGTCACCACTGTGTTTTCAGCTTCCCACT 1413
QY 1410 gacgtgctctacggcgctcaaccatgaattacacaaatgacaaatgaggtgagcttgcact 1469
DB 1414 GACCGTCCCTACGGCGCGCTCAACCATGAATTACAAAGTGGGATTTATAGGCGTTGCACT 1473
QY 1470 tgccttggtgcttgaactgggtcgtgcatgccaggaagcattatcaggggaccactt 1529
DB 1474 TGCTCTGTGTTCTTCAACTGGGTCTGTCATGCCAGGAACATTTATCAGGGACCCCACTT 1533
QY 1530 ggaagcttgacggcggtgctcgagacagaatttcaagttgggccaatga 1578
DB 1534 GGAGCTTGACGGACGGGTCTCGGAGCAGAATTTCAAGTTGGGCCATGA 1582

RESULT 3

AP000292/c 38209 bp DNA linear PRI 20-NOV-1999
LOCUS Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AML region,
DEFINITION clone:Q95D4, complete sequence.

ACCESSION AP000292
VERSION AP000292.1 GI:4835661
KEYWORDS HTG.

SOURCE Homo sapiens DNA, clone:Q95D4.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 38209)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.

TITLE Homo sapiens 38,209bp genomic DNA of 21q22.1

JOURNAL Published Only in DataBase (1999) In press

REFERENCE 2 (bases 1 to 38209)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.

TITLE Direct Submission

JOURNAL Submitted (13-MAY-1999) to the DDBJ/EMBL/GenBank databases.

Masahira Hattori, The Institute of Physical and Chemical Research

(RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1

Kitasato, Sagami-hara, Kanagawa 228-8555, Japan

(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,

TEL:81-42-778-9923, FAX:81-42-778-9924)

The sequence is a part of the data (ACCESSION No. AP000174 -

AP000194).

The sequencing project is supported by Japan Science Technology

Corporation (JST) and The Institute of Physical and Chemical

Research (RIKEN).

FEATURES Location/Qualifiers

source 1..38209

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="21"

/clone="Q95D4"

/map="21q22.1"

BASE COUNT 10658 a 8601 c 7890 g 11060 t

ORIGIN

Query Match 1.4%; Score 22; DB 9; Length 38209;

Best Local Similarity 100.0%; Pred. No. 1;


```

/evidence-not_experimental
repeat_region 12828..12963
/rpt_family="L1M4"
/evidence-not_experimental
repeat_region 13020..13319
/rpt_family="AluY"
/evidence-not_experimental
misc_feature 13154..13295
/note="GC score = 6.80 (142bp)"
/note="Region: GC content"
/evidence-not_experimental
repeat_region 13368..13480
/rpt_family="L1ME"
/evidence-not_experimental
repeat_region complement(13487..13851)
/rpt_family="MLT1E"
/evidence-not_experimental
repeat_region 13945..14002
/note="homology = 94.80%, score = 23, counts = 2"
/rpt_type=tandem
/rpt_unit-atgaaggcattaatatgtctgcaaac
/evidence-not_experimental
repeat_region 14514..14813
/rpt_family="AluJb"
/evidence-not_experimental
repeat_region 15160..15257
/rpt_family="FLAM_C"
/evidence-not_experimental
repeat_region 15320..15635
/rpt_family="L2"
/evidence-not_experimental
exon complement(15716..15811)
/note="Xpound exon prediction, score = 76% (0%)"
/evidence-not_experimental
repeat_region complement(15906..16328)
/rpt_family="MLT2FB"
/evidence-not_experimental
repeat_region 16333..16632
/rpt_family="L2"
/evidence-not_experimental
repeat_region complement(16858..17200)
/rpt_family="MLT1A2"
/evidence-not_experimental
repeat_region complement(17346..17473)
/rpt_family="MER20"
/evidence-not_experimental
repeat_region complement(17709..18006)
/rpt_family="AluSx"
/evidence-not_experimental
repeat_region complement(19099..19219)
/rpt_family="MER46"
/evidence-not_experimental
exon complement(19330..19431)
/note="GRAIL, score = 42.000%, comment = marginal"
/evidence-not_experimental
repeat_region complement(19809..20415)
/rpt_family="MER4A"
/evidence-not_experimental
exon 19969..20066
/note="MZF, score = 50.8%"
/evidence-not_experimental
repeat_region complement(20416..20828)
/rpt_family="MER57B"

```

Query Match 1.4%; Score 22; DB 9; Length 40697;
 Best Local Similarity 100.0%; Pred. No. 1;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1177 ggatgtatattcttaggttcta 1198
 |||||

Db 12042 GGATGTATATTCTAGGTTCTA 12021

```

RESULT 5
AP000043/c
LOCUS
DEFINITION
  Homo sapiens genomic DNA, chromosome 21q22.1, segment 14/28,
  complete sequence.
ACCESSION
  AP000043
VERSION
  AP000043.1 GI:3132353
KEYWORDS
  HTG.
SOURCE
  Homo sapiens DNA, clone:T1866-f90A8f_7.
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE
    1 (bases 1 to 100000)
    Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
    Homo sapiens genomic DNA, chromosome 21q
    Published Only in DataBase (1998) In press
  REFERENCE
    2 (bases 1 to 100000)
    Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
    Direct Submission
    Submitted (11-MAY-1998) to the DDBJ/EMBL/GenBank databases.
    Masahira Hattori, Kitasato University, Department of Science, JST
    Sequencing Laboratory: Kitasato 1-15-1, Sagami-hara 228, Japan
    (E-mail:hattori@hgc.ims.u-tokyo.ac.jp, Tel:0427-78-9732,
    Fax:0427-78-9561)
    This sequence is conducted by Kitasato University JST sequencing
    Laboratory as a JST sequencing team.
    Principal Investigator:Yoshiyuki Sakaki Ph.D.
    Phone: +81-3-5449-5622, Fax : +81-3-5449-5445,
    sakaki@hgc.ims.u-tokyo.ac.jp
    Sub-leader: Tadayoshi Shiba Ph.D., Masahira Hattori Ph.D. The
    sequence is submitted by:Human Genome Sequencing in ALIS project of
    JST
    Japan Science and Technology Corporation (JST)
    5-3, Yobancho, Chiyoda-ku, Tokyo 102-0028 Japan
    For further information about this sequence, including its location
    and relationship to other sequences, please visit our sequence
    archive Web site (http://www-alis.tokyo.jst.go.jp/HGS/top.html)
    or send email to webmaster@www-alis.tokyo.jst.go.jp'.
  FEATURES
    Location/Qualifiers
      1..100000
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="21"
        /clone="T1866-f90A8f_7"
        /map="21q22.1"
      BASE COUNT 27784 a 22015 c 21661 g 28540 t
      ORIGIN
        Query Match 1.4%; Score 22; DB 9; Length 100000;
        Best Local Similarity 100.0%; Pred. No. 0.91;
        Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
        Qy 1177 ggatgtatattcttaggttcta 1198
          |||||
        Db 30638 GGATGTATATTCTAGGTTCTA 30617

RESULT 6
AP000111/c
LOCUS
DEFINITION
  Homo sapiens genomic DNA of 21q22.1, GART and AML related,
  Q78C10-149C3 region, segment 14/20, complete sequence.
ACCESSION
  AP000111
VERSION
  AP000111.1 GI:4730845
KEYWORDS
  HTG.
SOURCE
  Homo sapiens DNA.
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE
    1 (sites)
    Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
    Homo sapiens 2,051,516bp genomic DNA of 21q22.1 GART and AML region

```


Gardiner, K., Nizetic, D., Francis, F., Lehrach, H., Reinhardt, R. and Yaspo, M. L.
 Direct Submission
 Submitted (10-APR-2000) to the DDBJ/EMBL/GenBank databases. The
 Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic
 Sciences Center, Human Genome Research Group * Institute of
 Molecular Biotechnology, Genome Analysis * Keio University School
 of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome
 Analysis * Max-Planck Institute for Molecular Genetics (addresses
 see below)
 On May 30, 2000 this sequence version replaced gi:7717330.
 The chromosome 21 mapping and sequencing consortium consisting of
 * RIKEN Genomic Sciences Center, Human Genome Research Group, *
 Sagami-hara 228-8555, Japan,
 * e.mail: hattori@gsc.riken.go.jp
 * URL: http://hgp.gsc.riken.go.jp/
 and
 * Institute of Molecular Biotechnology, Genome Analysis, *
 Beutenbergstrasse 11, D-07745 Jena, Germany,
 * e.mail: gscj-submit@genome.imb-jena.de
 * URL: http://genome.imb-jena.de/
 and
 * Keio University School of Medicine, Molecular Biology, * Tokyo
 160-8582, Japan,
 * e.mail: nshimizue@dm-b-med.keio.ac.jp
 * URL: http://www.dmb.med.keio.ac.jp/
 and
 * GBF, Dept. of Genome Analysis,
 * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
 info.genome@gbf.de
 * URL: http://genome.gbf.de/
 and
 * Max-Planck Institute for Molecular Genetics,
 * Innestrasse 73, D-14195 Berlin, Germany,
 * e.mail: info-chr21@molgen.mpg.de
 * URL: http://chr21.rz-berlin.mpg.de/
 AL163261: Submitted (10-APR-2000).

FEATURES
 source
 1..340000
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 <1..16929
 /note="Accession No. AP000288"
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q22.1"
 /clone="RQ57C2, 5' partial"
 /clone_lib="CNP21 pl library"
 complement(722..1108)
 /note="MTIB"
 /rpt_family="LTR/MaLR"
 /rpt_type=DISPERSED
 complement(1428..1508)
 /note="MIR"
 /rpt_family="SINE/MIR"
 /rpt_type=DISPERSED
 1794..1944
 /note="MIR"
 /rpt_family="SINE/MIR"
 /rpt_type=DISPERSED
 complement(3244..3344)
 /note="MTIC"
 /rpt_family="LTR/MaLR"
 /rpt_type=DISPERSED
 complement(3634..3765)
 /note="MTIB"
 /rpt_family="LTR/MaLR"
 /rpt_type=DISPERSED
 4459..4598
 /note="AluJo"
 /rpt_family="SINE/Alu"

repeat_region
 complement(4657..4725)
 /note="L2"
 /rpt_family="LINE/L2"
 /rpt_type=DISPERSED
 complement(4753..5044)
 /note="L2"
 /rpt_family="LINE/L2"
 /rpt_type=DISPERSED
 5033..5113
 /note="(TC)n"
 /rpt_family="Simple_repeat"
 /rpt_type=TANDEM
 complement(5181..5289)
 /note="L2"
 /rpt_family="LINE/L2"
 /rpt_type=DISPERSED
 complement(5522..5636)
 /note="MER15"
 /rpt_family="DNA"
 /rpt_type=DISPERSED
 7429..7517
 /note="MIR"
 /rpt_family="SINE/MIR"
 /rpt_type=DISPERSED
 8064..8216
 /note="MER5A"
 /rpt_family="DNA/MER1_type"
 /rpt_type=DISPERSED
 complement(8488..8870)
 /note="L2"
 /rpt_family="LINE/L2"
 /rpt_type=DISPERSED
 9221..9278
 /note="MER5A"
 /rpt_family="DNA/MER1_type"
 /rpt_type=DISPERSED
 9279..9452
 /note="FAM"
 /rpt_family="SINE/Alu"
 /rpt_type=DISPERSED
 9453..9525
 /note="MER5A"
 /rpt_family="DNA/MER1_type"
 /rpt_type=DISPERSED
 complement(9644..9789)
 /note="MIR"
 /rpt_family="SINE/MIR"
 /rpt_type=DISPERSED
 complement(9819..10148)
 /note="MTIC"
 /rpt_family="LTR/MaLR"
 /rpt_type=DISPERSED
 10149..10454
 /note="L1PA5"
 /rpt_family="LINE/L1"
 /rpt_type=DISPERSED
 complement(10455..10617)
 /note="MTIC"
 /rpt_family="LTR/MaLR"
 /rpt_type=DISPERSED
 complement(10661..10844)
 /note="MIR"
 /rpt_family="SINE/MIR"
 /rpt_type=DISPERSED
 10921..10953
 /note="(TTCC)n"
 /rpt_family="Simple_repeat"
 /rpt_type=TANDEM
 complement(10954..11263)
 /note="AluSg"
 /rpt_family="SINE/Alu"
 /rpt_type=DISPERSED

```

repeat_region 11567..11608
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
repeat_region 11574..11618
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
repeat_region complement(11625..11680)
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
repeat_region complement(11681..11978)
/note="AluJb"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region complement(11979..12068)
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
repeat_region 13530..13835
/note="AluXs"
/rpt_family="SINE/Alu"

Query Match 1.48; Score 22; DB 9; Length 340000;
Best Local Similarity 100.0%; Pred. No. 0.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1177 ggatgatattttaggttcta 1198
DB 156788 CGATGATATTTAGGTCTA 156767

RESULT 9
HSPACTIXG/c HSPACTIXG 11873 bp DNA linear PAT 02-MAR-1995
LOCUS H.sapiens factor IX gene.
DEFINITION H.sapiens factor IX gene.
ACCESSION A22491
VERSION A22491.1 GI:825659
KEYWORDS human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 11873)
AUTHORS Brownlee,G.G. and Choo,K.H.
TITLE Molecular cloning of the gene for human anti-haemophilic factor IX
JOURNAL Patent: EP 0107278-A 13 02-MAY-1984;
NATIONAL RESEARCH DEVELOPMENT CORPORATION
FEATURES
source 1..11873
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 3617 a 2081 c 2165 g 3180 t 830 others
ORIGIN

Query Match 1.38; Score 21; DB 6; Length 11873;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 827 cagtggtcagtcgaatcataa 847
DB 1286 CAGTGGTCAGTGAATCATAA 1266

RESULT 10
HUMFIXG/c HUMFIXG 38059 bp DNA linear PRI 30-APR-1996
LOCUS Human coagulation factor IX gene, complete cds.
DEFINITION Human coagulation factor IX gene, complete cds.
ACCESSION K02402
VERSION K02402.1 GI:182612
KEYWORDS Alu repeat; Christmas factor; KpnI repetitive sequence;

```

```

SOURCE
antihemophilic factor B; factor IX; repeat region; simple
repetitive sequence.
Homo sapiens (clone: FIX-lambda-[6,36,53,61].) (tissue library:
T.Maniatis et al.) DNA; and Homo sapiens (clone: FIX-lambda-4243)
DNA.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 38059)
AUTHORS Yoshitake,S., Schach,B.G., Foster,D.C., Davie,E.W. and Kurachi,K.
TITLE Nucleotide sequence of the gene for human factor IX (antihemophilic
factor B)
JOURNAL Biochemistry 24 (14), 3736-3750 (1985)
MEDLINE 86000558
REFERENCE 2 (bases 23487 to 23556)
AUTHORS Rees,D.J., Rizza,C.R. and Brownlee,G.G.
TITLE Haemophilia B caused by a point mutation in a donor splice junction
of the human factor IX gene
JOURNAL Nature 316 (6029), 643-645 (1985)
MEDLINE 85296286
REFERENCE 3 (bases 23378 to 23387)
AUTHORS Graham,J.B., Lubahn,D.B., Lord,S.T., Kirshtein,J., Nilsson,I.M.,
Wallmark,A., Ljung,R., Frazier,L.D., Ware,J.L., Lin,S.W.,
Stafford,D.W. and Bosco,J.
TITLE The Maimo polymorphism of coagulation factor IX, an immunologic
polymorphism due to dimorphism of residue 148 that is in linkage
disequilibrium with two other F.IX polymorphisms
JOURNAL Am. J. Hum. Genet. 42 (4), 573-580 (1988)
MEDLINE 88161064
REFERENCE 4 (sites)
AUTHORS Hirosewa,S., Fahner,J.B., Sallier,J.-P., Wu,C.-T., Lovrien,E. and
Kurachi,K.
TITLE Structural and functional basis of the developmental regulation of
human factor IX gene: factor IX Leyden
JOURNAL Unpublished (1990)
COMMENT [1] kindly submitted on floppy by K.Kurachi,
05-AUG-1985.
[1] notes a potential TATA box (2939-2942) and polydenylation
signal (35701-35706); and notes two start codons (downstream of the
start codon annotated below) that may be alternative and/or
preferred starts for the factor IX prepropeptide. Several tracts
of simple repetitive sequence are present [1], including regions
with the potential for hairpin and/or Z-DNA formation. [1]
describes six long open reading frames in the intron and on the
complementary strand.
FEATURES
source Location/Qualifiers
1..38059
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="Xq26.3-q27.1"
/clone="FIX-lambda-4243"
/clone="FIX-lambda-[6,36,53,61]."
/cell_line="49, XXXXX"
/cell_type="fibroblast"
/tissue_lib="T.Maniatis et al."
repeat_region 1..2596
/rpt_family="KpnI"
prim_transcript 2966..35722
/gene="F9"
/note="FIX mRNA"
2966..35722
/gene="F9"
2966..3082
/gene="F9"
/number=1
CDS
1018(2995..3082,9291..9454,9643..9667,13357..13470,
20634..20762,23328..23530,33004..33118,33787..34334)
/gene="F9"
/note="precursor"
/codon_start=1
/product="factor IX"
/protein_id="AAB59620.1"

```



```
/db_xref="GI:182613"
/db_xref="GDB:G00-119-900"
/translation="MQRVNMIMASPGILITICILLYLSARCTVFLDHNANKILNRP
KRNKGLEEFVQGLRECEKSCFEAREVFENTERFERWKQYVDGQCESNPC
LNGSKDDINSIECWCPFGEKNCEDVTCNKNGRCQFCNKSADNKKVCSCTEG
YRLAENKSCPEAVFPFGCRVSVQSKLTRAFAVFPDVIYVNSTEATILDRITQST
QSFDFRVGGEDAKPGQFPQVVLNGKVDFAFCGGSIVNEKWIIVTAACHVETGVKIT
VWAGHNIETETHEQRKRVIRIIPHNNYNAIINKYHDIALLLEDEPLVINSYVTFI
CIADREYNILFKFGSYVSGWRVFKHGRSALVQLRVPLVDRAICLRSKFTIYN
NMFCAGFHEGGDSCGDSGGPHVTEVGTSELTGIIISWGECCAMKGYGIYTKVSRV
VNWIKETKLT"
join(2905..3082,9291..9340)
/genes="F9"
/number=5
/note="G00-119-900"
/replace="a"
23387
/genes="F9"
/number=1
/note="G00-119-900"
/replace="a"
7298..7593
/rpt_family="Alu"
8469..8520
/genes="F9"
/note="t...50 bp...a in FIX-lambda-4243; ta in
FIX-lambda-[36,61]; G00-119-900"
/citation=[1]
/replace="ta"
9291..9454
/genes="F9"
/note="G00-119-900"
/number=2
join(9341..9454,9643..9667,13357..13470,20634..20762,
23328..23360)
/genes="F9"
/note="G00-119-900"
/product="factor IX light chain"
9455..9642
/genes="F9"
/note="G00-119-900"
/number=2
9643..9667
/genes="F9"
/note="G00-119-900"
/number=3
9668..13356
/genes="F9"
/note="G00-119-900"
/number=3
10041
/genes="F9"
/note="g in one allele; c in another allele (loss of XmnI
recognition pattern); G00-119-900"
/replace="c"
13357..13470
/genes="F9"
/note="G00-119-900"
/number=4
13471..20633
/genes="F9"
/note="G00-119-900"
/number=4
14076..14079
/genes="F9"
/note="tcga in one allele; nnnn in another allele (loss of
TagI site); G00-119-900"
/replace="nnnn"
18165..20265
/rpt_family="KpnI"
20416
/genes="F9"
/citation=[4]
/replace="a"
20634..20762
/genes="F9"
```

```

/note="G00-119-900"
/number=5
20763..23327
/genes="F9"
/note="G00-119-900"
/number=5
23328..23530
/genes="F9"
/note="G00-119-900"
/number=6
23387
/genes="F9"
/note="g in [1]; a in [3] Ala->Thr; G00-119-900"
/replace="a"
23387
/genes="F9"
/note="g in one allele; a in another allele"
/replace="a"
join(23486..23530,33004..33118,33787..34331)
/genes="F9"
/note="G00-119-900"
/product="factor IX heavy chain"
23531..33003
/genes="F9"
/note="G00-119-900"
/number=6
23531
/genes="F9"
/note="g in [1]; t in [2] (haemophilia patient);
G00-119-900"
/replace="t"
24172..24475
/rpt_family="Alu"
25863..26091
/rpt_family="Alu"
31537..31809
/rpt_family="Alu"
33004..33118
/genes="F9"
/note="G00-119-900"
/number=7
33119..33786
/genes="F9"
/note="G00-119-900"
/number=7
33787..35722
/genes="F9"
/note="G00-119-900"
/number=8

Query Match 1.3%; Score 21; DB 9; Length 38059;
Best Local Similarity 100.0%; Pred.No.3.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 827 cagtggtcagtcgaatcataa 847
|||||
Db 17480 CAGTGGTCAGTCGATCATAA 17460

RESULT 11
AL645665/c
LOCUS
DEFINITION Homo sapiens chromosome X clone RP11-963P9, *** SEQUENCING IN
PROGRESS ***, 3 unordered pieces.
ACCESSION AL645665
VERSION AL645665.9 GI:18491387
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEPIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Heath, P.
```

TITLE Direct Submission
JOURNAL Submitted (31-JAN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT On Feb 4, 2002 this sequence version replaced gi:18476880.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: BA963P9
 ----- Summary Statistics
 Sequencing program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: dye-terminator Big Dye; 100% of reads
 Consensus quality: 152173 bases at least Q40
 Consensus quality: 152586 bases at least Q30
 Consensus quality: 152784 bases at least Q20
 Insert size: 153415; sum-of-contigs
 Insert size: 144208; 38.0% error; agarose-fp
 Quality coverage: 20.53x in Q20 bases; sum-of-contigs Quality coverage: 22.15x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
 * 1 28734: contig of 28734 bp in length
 * 28735 28834: gap of 100 bp
 * 28835 44671: contig of 15837 bp in length
 * 44672 44771: gap of 100 bp
 * 44772 153615: contig of 108844 bp in length.

FEATURES
 source
 1. .153615
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /clone="Rp11-963P9"
 /clone_lib="RPC1-11.4"
 1. 28734
 /note="assembly_fragment:06384
 fragment_chain:1
 clone_end:SP6
 vector_side:left"
 28835..44671
 /note="assembly_fragment:02107
 fragment_chain:1"
 44772..153615
 /note="assembly_fragment:00800
 fragment_chain:1"
 50474 a 29730 c 28392 g 44819 t 200 others
BASE COUNT
ORIGIN

Query Match 1.3%; Score 21; DB 2; Length 153615;
 Best Local Similarity 100.0%; Pred. No. 3.3;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 827 cagtggctcagtgaatcataa 847
 Db 93977 CAGTGGTCAGTGAATCATAA 93957
 |||||
 RESULT 12
 HS88D7/c 158557 bp DNA linear PRI 23-NOV-1999
 LOCUS Human DNA sequence from clone 88D7 on chromosome Xq25-26.3 Contains
 DEFINITION F9 (coagulation factor IX (plasma thromboplastic component,

ACCESSION AL033403.1
VERSION AL033403.1
KEYWORDS HTG; Christmas factor; db1 oncogene; F9; factor IX; glycoprotein; oncogene; phosphoprotein; proto-oncogene.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 158557)
TITLE Bird.C.
JOURNAL Direct Submission
COMMENT Submitted (09-NOV-1998) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence is the entire insert of clone 88D7.
 This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
 This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/ChrX
 88D7 is from the library RPC16 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/ VECTOR: pPAC4.
 Location/Qualifiers
 1. .158557
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="q25-26.3"
 /clone="RP6-88D7"
 /clone_lib="RPC1-6"
 2. .3009
 /note="L1PA2 repeat: matches 3136..5144 of consensus"
 3023..33684
 /note="L1M4 repeat: matches 4639..5306 of consensus"
 3693..3775
 /note="L1MC/D repeat: matches 5517..5601 of consensus"
 join(<4083..4171,10329..10492,10681..10705,14395..14508,21682..21810,24381..24583,34060..34174,34843..36777)
 /gene="F9"
 /note="match: DNA M21757 J00137 A13957 M33826 M11309 J00136 A22493 M23109 M35672 U51135 M26247 M26233 M26234 M26235 M26236 M26237 Z50155; match: STS G28615"
 /evidence=not_experimental
 4083..36777
 /gene="F9"
 join(4099..4171,10329..10492,10681..10705,14395..14508,21682..21810,24381..24583,34060..34174,34843..35390)
 /gene="F9"
 /note="match: protein P00740 P19540 P16294 P00741 P16293 P16296 P16295 P16292 P16291 P25155"
 /codon_start=1
 /evidence=not_experimental
 /product="F9 (coagulation factor IX (plasma thromboplastic component, Christmas disease, haemophilia B))"
 /protein_id="CAA21954.1"
 /db_xref="GI:3859055"
 /db_xref="SPTREMBL:O14316"
 /translation="MIMASPGITICLLCYLLSACTVFLDHENANKILNRPKRYNS
 GKLEFVQGNLERECMEKCSFEAREVFENTERTFWKQYVDGDCSNPCINGSS

Christmas disease, haemophilia B)), db1 oncogene. EST, STS, GSS, complete sequence.
 AL033403
 HTG; Christmas factor; db1 oncogene; F9; factor IX; glycoprotein; oncogene; phosphoprotein; proto-oncogene.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 158557)
 Bird.C.
 Direct Submission
 Submitted (09-NOV-1998) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence is the entire insert of clone 88D7.
 This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
 This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/ChrX
 88D7 is from the library RPC16 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/ VECTOR: pPAC4.
 Location/Qualifiers
 1. .158557
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="q25-26.3"
 /clone="RP6-88D7"
 /clone_lib="RPC1-6"
 2. .3009
 /note="L1PA2 repeat: matches 3136..5144 of consensus"
 3023..33684
 /note="L1M4 repeat: matches 4639..5306 of consensus"
 3693..3775
 /note="L1MC/D repeat: matches 5517..5601 of consensus"
 join(<4083..4171,10329..10492,10681..10705,14395..14508,21682..21810,24381..24583,34060..34174,34843..36777)
 /gene="F9"
 /note="match: DNA M21757 J00137 A13957 M33826 M11309 J00136 A22493 M23109 M35672 U51135 M26247 M26233 M26234 M26235 M26236 M26237 Z50155; match: STS G28615"
 /evidence=not_experimental
 4083..36777
 /gene="F9"
 join(4099..4171,10329..10492,10681..10705,14395..14508,21682..21810,24381..24583,34060..34174,34843..35390)
 /gene="F9"
 /note="match: protein P00740 P19540 P16294 P00741 P16293 P16296 P16295 P16292 P16291 P25155"
 /codon_start=1
 /evidence=not_experimental
 /product="F9 (coagulation factor IX (plasma thromboplastic component, Christmas disease, haemophilia B))"
 /protein_id="CAA21954.1"
 /db_xref="GI:3859055"
 /db_xref="SPTREMBL:O14316"
 /translation="MIMASPGITICLLCYLLSACTVFLDHENANKILNRPKRYNS
 GKLEFVQGNLERECMEKCSFEAREVFENTERTFWKQYVDGDCSNPCINGSS

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalish, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 22 Row: 1 Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Genomescan gene prediction, similarity but not identity
to protein.

FEATURES Location/Qualifiers
SOURCE 1..2879
 /organism="Homo sapiens"
 /db_xref="LocusID:5211"
 /db_xref="taxon:9606"
 /clone="MGC:15386 IMAGE:3347301"
 /tissue_type="Kidney, renal cell adenocarcinoma"
 /clone_lib="NIH_MGC_14"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"
 16..2358
 /codon_start=1
 /product="phosphofructokinase, liver"
 /protein_id="AA07536.1"
 /db_xref="GI:14043101"
 /translation="MAAVDLKLRASGAGKAIGVLTSGDQAGNNAVAVTRMGIVV
GAKFLIYEGYELGVEGGENIKQANWLSVNIIOGGTIIIGSARCAFTTREGRRAA
YNLVQGITNLGCVGGDSLTCANIFRSENGSLLEELVAECKISETTARTYSHLNAG
LVGSINDPCGDTMTIGTDSALHRIMEVIDAITTAQSHQRTFVLEVMGRHCYLAIV
SALASADWLFIPPEAPDGEWENFMCERLGTFRSGRLNIIIEAGAIIDRNGKPISS
SVYKDLVORLGFDTVTVLGHVORGGTSPAFDRILSKMGMEAVMALLEATPDTPAC
VYTLSCNOSVRLPMECVQMTKEVOKANDKREDEATOLGSGSEFENNINLYKLAHOK
PKKSNFSLALINVGAPAGNNAVRSAYRTGISHGTYVYVHDGFEGLAKGVQVEY
GHDVAGWLRGSGSLMGTKRLTPRGQESIVENIRIYGIHALLVVGFEAYEGYQLV
EARGRYEELICVMVIPATISNNVPTDPSIGSDTAVNAAMESCDRIKQASGTRRV
FIVETMGGYCGYLATVTGIAVGADAAVYFDPFNHDLKVNVEHMTKMKRTDIOGLV
LRNEKCHDYTTTFEFLYNYLSSGKGVFDCRTNVLGHQQGAGPTPFDRTNYTKLGVKA
MLWLSKRLREVRYKRVFANAPDSACVIGLAKKAVAFSPVTELKKTDFEHRMPREQM
WLSLRMLAKLAQYRISMAAYVSGELEHVRTRTLMDKGF"

BASE COUNT 554 a 875 c 932 g 518 t
ORIGIN

Query Match 1.3%; Score 20; DB 9; Length 2879;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 505 gtctctctcatctacgaggg 524
 |||||
Db 157 GTCTTCTCATCTACGAGGG 176

Search completed: April 27, 2002, 05:44:33
Job time: 17910 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:57:41 ; Search time 114.27 Seconds
(without alignments)
737.276 Million cell updates/sec

Title: US-09-882-694A-5
Perfect score: 2549
Sequence: 1 NVLSPDEYKSELFNFVFS.....EEGLKAYLEPKTINHVNIE 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_19.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-----------|---------------------|
| 1 | 889.5 | 34.9 | 517 | 4 Q9BV45 | Q9bv45 homo sapien |
| 2 | 889 | 34.9 | 497 | 3 Q9C1Q6 | Q9clq6 emericeila |
| 3 | 888 | 34.8 | 497 | 3 Q9C1Q5 | Q9clq5 emericeila |
| 4 | 883 | 34.6 | 497 | 3 Q9C1Q4 | Q9clq4 emericeila |
| 5 | 882.5 | 34.6 | 519 | 11 Q9CZS1 | Q9czs1 mus musculus |
| 6 | 878.5 | 34.5 | 501 | 11 Q35945 | Q35945 mus musculus |
| 7 | 872 | 34.2 | 497 | 3 Q9P303 | Q9p303 cladosporiu |
| 8 | 853.5 | 33.5 | 537 | 5 Q9TXM0 | Q9txm0 caenorhabdi |
| 9 | 849.5 | 33.3 | 518 | 13 Q9OXS8 | Q9oxs8 brachydanio |
| 10 | 845.5 | 33.2 | 518 | 13 Q9DEX5 | Q9dex5 xenopus lae |
| 11 | 842.5 | 33.1 | 518 | 13 Q9OY03 | Q9oy03 brachydanio |
| 12 | 826 | 32.4 | 502 | 13 Q9PWJ3 | Q9pwj3 xenopus lae |
| 13 | 825 | 32.4 | 517 | 13 Q9I8W8 | Q9i8w8 poephila gu |
| 14 | 822 | 32.2 | 502 | 13 Q9YGY2 | Q9ygy2 xenopus lae |
| 15 | 820 | 32.2 | 503 | 3 O14293 | O14293 schizosacch |
| 16 | 818.5 | 32.1 | 512 | 11 Q9JHW9 | Q9jhw9 mus musculus |

| | | | | | |
|----|-------|------|-----|-----------|---------------------|
| 17 | 813.5 | 31.9 | 512 | 11 Q9BQP7 | Q9bqp7 mus musculu |
| 18 | 811.5 | 31.8 | 497 | 3 Q92460 | Q92460 ustilago ma |
| 19 | 809.5 | 31.8 | 492 | 5 Q9NJH2 | Q9njh2 placopecten |
| 20 | 809.5 | 31.8 | 512 | 11 Q9JI72 | Q9ji72 mus musculu |
| 21 | 807.5 | 31.7 | 510 | 5 Q20780 | Q20780 caenorhabdi |
| 22 | 793.5 | 31.1 | 512 | 13 Q9DD46 | Q9dd46 gallus galli |
| 23 | 793 | 31.1 | 443 | 11 Q9I2D7 | Q9i2d7 rattus norv |
| 24 | 784 | 30.8 | 507 | 10 Q94JC5 | Q94jc5 oryza sativ |
| 25 | 781.5 | 30.7 | 501 | 10 Q9LV57 | Q9lv57 arabidopsis |
| 26 | 781 | 30.6 | 538 | 10 Q9SU63 | Q9su63 arabidopsis |
| 27 | 779 | 30.6 | 494 | 2 Q33455 | Q33455 pseudomonas |
| 28 | 779 | 30.6 | 549 | 10 Q43274 | Q43274 zea mays (m |
| 29 | 776.5 | 30.5 | 520 | 5 Q9VLC5 | Q9vlc5 drosophila |
| 30 | 773.5 | 30.3 | 496 | 3 Q9URW9 | Q9urw9 schizosacch |
| 31 | 773 | 30.3 | 549 | 10 Q9AG64 | Q9ag64 zea mays (m |
| 32 | 772.5 | 30.3 | 422 | 4 Q9UBR8 | Q9ubr8 homo sapien |
| 33 | 772.5 | 30.3 | 500 | 2 Q9F3V7 | Q9f3v7 pseudonocar |
| 34 | 772.5 | 30.3 | 659 | 5 Q9VB96 | Q9vb96 drosophila |
| 35 | 763 | 29.9 | 502 | 10 Q9LRE9 | Q9lre9 oryza sativ |
| 36 | 761 | 29.9 | 549 | 10 Q93X16 | Q93x16 hordeum vul |
| 37 | 760 | 29.8 | 549 | 10 Q9FRX7 | Q9frx7 oryza sativ |
| 38 | 759 | 29.8 | 549 | 10 Q9LLR2 | Q9llr2 oryza sativ |
| 39 | 751 | 29.5 | 542 | 10 P93344 | P93344 nicotiana t |
| 40 | 749.5 | 29.4 | 414 | 13 Q9W617 | Q9w617 xenopus lae |
| 41 | 749 | 29.4 | 518 | 3 Q12648 | Q12648 pichia angu |
| 42 | 741 | 29.1 | 534 | 10 Q9AC67 | Q9ac67 arabidopsis |
| 43 | 737.5 | 28.9 | 495 | 16 Q34660 | Q34660 bacillus su |
| 44 | 736 | 28.9 | 553 | 10 Q9LR16 | Q9lr16 oryza sativ |
| 45 | 728.5 | 28.6 | 519 | 10 Q9ZUB6 | Q9zub6 arabidopsis |

ALIGNMENTS

RESULT 1

ID Q9BV45 PRELIMINARY; PRT; 517 AA.

AC Q9BV45; 2001 (TREMREL. 17, Created)

DT 01-JUN-2001 (TREMREL. 17, Last sequence update)

DE 01-DEC-2001 (TREMREL. 19, Last annotation update)

DE SIMILAR TO ALDEHYDE DEHYDROGENASE 5.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=EYE, RETINOBLASTOMA;

RA Strausberg R.;

RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC001619; AAH01619.1; -

DR HSSP; P05091; ICW3

DR InterPro; IPR002086; Aldehyde_dehydr.

DR Pfam; PF00171; algedh; 1.

DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.

DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.

SQ SEQUENCE 517 AA; 57238 MW; A628E448D1E8689D CRC64;

Query Match 34.9%; Score 889.5; DB 4; Length 517;

Best local similarity 42.0%; Pred. No. 8e-61;

Mismatches 178; Indels 33; Gaps 12;

Matches 210; Conservative 79;

Qy 2 VLSPEYKSELFNFVSSKGSERLTLTPNPDESTVATDVHVA--NAADVSAVAASVQ 59

Db 29 ILNPDIPNQLFINNEQDAVSKKTFPTVNTTGEVIG---HVAEGDRAVDRAVKAARE 85

Qy 60 AVKKG-PWKKTTGQRAACMLKFAADLAEKNAEKLALESPTGRVSMITHFDIPNWSV 118

Db 86 AFRLGSPWRMDASERGLNLLADLVERDVRVYLASLETLDNGKPFQESYALDLDEIVK 145

Qy 119 FRYVAGNADKTAGTFFPDNCKPNW-RYEPWGVAGIASNATFLYVGVKIAFALAAGCS 177

```
Db 146 YRYFAGWADKWHGKTIIPMDQHFCTRRHEPVGVCGQIIPWNPFLVMOGKWLAPALATGNT 205
QY 178 FIFKASEKSPGLVGLAPLFAAGPPGVVQFLTGARVT-GEALASHMDIAKISFTRSVG 236
Db 206 VVKVAEOPTLSALYLASLIKAGEPPGVVNIITGYGPTAGAAIAOHMDVDVKAFTGST 265
QY 237 GGRAVKATLKNRVRTLELGEK-PTIVFNEAPLE- --RQSGESAKDFSKFGQIWWPPS 292
Db 266 VGHILQKAAGDNLKRVTLLELGKSPSIVLADADMEHAEVQCHEAL- --PFNMGQCCACC 323
QY 293 CLLVQW- -----NLAERFHVHSGFGCGCORWLGONPLKPRTHGPPVDSQVDRVLG 345
Db 324 RTFVEESIYNFLERTVEK- -----AKORKVG-NPFELDTQGGPOVDKEQFERVLG 372
QY 346 NIDVG-KDTAQLLTGVGRKDGKFALEPTIFVNPKPCKSWFEEIFGPVLSIKTFKTEE 404
Db 373 YIOLGOKGAKLLCGGERGEGFFIKPTVFGVQDDMRKAEIEFGPVQPLFKFKTEE 432
QY 405 AIEIANDTYGLASVIYTKSLNRGLRVSSALETGGVSINFPFIPETQTPFGCKMGSGGR 464
Db 433 VVERANTRYGLAAAVFTRDLQKAMFTQALQAGTWWNTYINVTCHTPFGGFKESGNGR 492
QY 465 ELGEGKLKAYLEPKTINIHV 484
Db 493 ELGEDGLKAYTEVKIVTKV 512
```

```
RESULT 2
Q9C106
ID Q9C106 PRELIMINARY: PRT: 497 AA.
AC Q9C106;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE ALDEHYDE DEHYDROGENASE ALDH.
GN ALDA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21269214; PubMed=11102439;
RA Filippi M., Mathieu M., Cirpus I., Panozzo C., Fellenbok B.;
RT "Regulation of the aldehyde dehydrogenase gene (aldA) and its role in
RT the control of the co-inducer level necessary for induction of the
RT ethanol utilization pathway in Aspergillus nidulans.";
RL J. Biol. Chem. 276:6950-6958(2001).
DR EMBL; AF260123; AAK18072.1; -.
DR HSSP; P05091; ICW3.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
FT VARIANT 191 191 E -> Q.
FT VARIANT 208 208 G -> P.
FT VARIANT 247 247 R -> P.
FT VARIANT 409 409 E -> V.
SQ SEQUENCE 497 AA; 54138 MW; 9C6CFC62731A25BD CRC64;
```

```
Query Match 34.9%; Score 889; DB 3; Length 497;
Best Local Similarity 43.1%; Pred. No. 8.3e-61;
Matches 207; Conservative 78; Mismatches 173; Indels 22; Gaps 11;

QY 12 LFINNEFVSSKGSERLTLTNPWDESTVATDVHVAADVDVAASVQAVKGPWKFTG 71
Db 21 LFINNEFVSGEKTQFVINPSNEK-VITSVHEATEKDVDVAVAAA-RAAFEGPWRQVTP 78
QY 72 AORAAKMLKFDALAEKNAEKLARLESPTGRPVSMITHFDIPNMYSVFRYAGWADKTAG 131
Db 21 LFINNEFVSGEKTQFVINPSNEK-VITSVHEATEKDVDVAVAAA-RAAFEGPWRQVTP 78
QY 72 AORAAKMLKFDALAEKNAEKLARLESPTGRPVSMITHFDIPNMYSVFRYAGWADKTAG 131
Db 79 SERGLINKLADLMERDITLAAIESLNGKAFTH-AKVDLANSIGCLRYRYAGWADKTAG 137
QY 72 AORAAKMLKFDALAEKNAEKLARLESPTGRPVSMITHFDIPNMYSVFRYAGWADKTAG 131
Db 79 SERGLINKLADLMERDITLAAIESLNGKAFTH-AKVDLANSIGCLRYRYAGWADKTAG 137
```

```
QY 132 KTFPEDNCKPNW-RYEPMGVCAGIASWNATLYVGWKIAPALAAAGSPFKASEKSPLCV 190
Db 138 QTIDTNPETLTTRHEPVGVCGQIIPWNPFLVMOGKWLAPALATGNT 197
QY 191 LGLAPLFAAGPPGVVQFLTG-ARVTGEPALASHMDIAKISFTRSVGGRAVKQATLKN 249
Db 198 LYAAKLIKAGEPPAGVINISGFRGTAGAAISSHMDIDKVAFTGSTVLGRTILQAAAKSN 257
QY 250 MKRVTLLELGEK-PTIVFNEAPLE- --RQSGESAKDFSKFGQIWWPPSCLLVQWGNLAEKF- 305
Db 258 LAKVTLLELGGKSPNIVEDDADIDNAISMANFGIFTHGQCCAGSRILVQEG- IYDKFVA 316
QY 306 ---HGVRHSGFGCGCORWLGONPLKPRTHGPFVDSQVDRVLGNTDVGKDT-AQLLTGVG 361
Db 317 RFKERAQKNKVG- -----NPFEDTFOGPOVSQLOFDRIMEYINHKKAGATVATGCD 368
QY 362 RKDGKGALEPTIFVNPKPCKSWFEEIFGPVLSIKTFKTEEAEIANDTYGLASVIY 421
Db 369 RHNGEGYFQPTVFTVDTSDMKIAQEEIFGPVVTQKFKDEABAIKIGNSTDYGLAAAVH 428
QY 422 TKSLNRGLRVSSALETGGVSINFPFIPETQTPFGCKMGSGRELGEGLKAYLEPKTIN 481
Db 429 TKNVNTAIRVSNALKAGTWINNKNISQAPFGGKQSGGLGRELGSYALENTQIKTVH 488

RESULT 3
Q9C105
ID Q9C105 PRELIMINARY: PRT: 497 AA.
AC Q9C105;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE ALDEHYDE DEHYDROGENASE ALDH15.
GN ALDA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21269214; PubMed=11102439;
RA Filippi M., Mathieu M., Cirpus I., Panozzo C., Fellenbok B.;
RT "Regulation of the aldehyde dehydrogenase gene (aldA) and its role in
RT the control of the co-inducer level necessary for induction of the
RT ethanol utilization pathway in Aspergillus nidulans.";
RL J. Biol. Chem. 276:6950-6958(2001).
DR EMBL; AF260124; AAK18073.1; -.
DR HSSP; P05091; ICW3.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
FT VARIANT 286 286 V -> A.
SQ SEQUENCE 497 AA; 54166 MW; 65849D1DCA94A5F5 CRC64;
```

```
Query Match 34.8%; Score 888; DB 3; Length 497;
Best Local Similarity 43.1%; Pred. No. 9.9e-61;
Matches 207; Conservative 78; Mismatches 173; Indels 22; Gaps 11;

QY 12 LFINNEFVSSKGSERLTLTNPWDESTVATDVHVAADVDVAASVQAVKGPWKFTG 71
Db 21 LFINNEFVSGEKTQFVINPSNEK-VITSVHEATEKDVDVAVAAA-RAAFEGPWRQVTP 78
QY 72 AORAAKMLKFDALAEKNAEKLARLESPTGRPVSMITHFDIPNMYSVFRYAGWADKTAG 131
Db 79 SERGLINKLADLMERDITLAAIESLNGKAFTH-AKVDLANSIGCLRYRYAGWADKTAG 137
QY 132 KTFPEDNCKPNW-RYEPMGVCAGIASWNATLYVGWKIAPALAAAGSPFKASEKSPLCV 190
Db 138 QTIDTNPETLTTRHEPVGVCGQIIPWNPFLVMOGKWLAPALATGNT 197
```



```

QY 191 LGLAPFAEAGFPVGVQFTLG-ARVTGEALASHMDIAKISFTRSVGGGAVKQATLAKSN 249
DB 198 LYAAKLKEAGFPAGVINVISGFRAGAAISHMDIDKRVATGTLVGRITLIQAAKSN 257
QY 250 MKRVTLEIGEK-PTIVNEAPLERQ-SGESAKDFSKFGQIWWPPSCLLVQWGNLAEKF-- 305
DB 258 LKVTLELGGKSPNIVFDDADIDNAISWNFHGCCAGSRILVQEG-YYDKFVA 316
QY 306 ---HGVRHGSGFGCCQRLWGNLQONLEPKRTHGPFVDRKSOYDRVLGNIDVCKDT-AOLLTGVG 361
DB 317 REKERAOKNVG-----NPFQDTFQGPQVSOLOFDRIMEYINHKKAGATVATGSD 368
QY 362 RKDGKFAIEPTTFVNPCKSGKIWFEEIFGPVLSIKTFKTEETETIANDTDTYGLASVIY 421
DB 369 RHGNEGYFIQPTFTDVTSDMKIAQEEIFGPVVTIQKFDEAEAIKIGNSTIDYGLAAVH 428
QY 422 TKSINRGLRVSSALETGGVSNFPFIPETQTPFGMKQSGSGRELGEGLKAYLEPKTIN 481
DB 429 TKNVNTAIRVSNALKAGTVINNYNMISYQAPFGGKQSGLGRGSLGVALENTYQIKTVH 488

RESULT 4
Q9C1Q4 PRELIMINARY; PRT; 497 AA.
AC Q9C1Q4;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE ALDEHYDE DEHYDROGENASE ALDH57.
GN ALDA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
RX NCBI_TaxID=5072;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21269214; PubMed=11102439;
RA Flippin M., Mathieu M., Cirpus I., Panozzo C., Fellenbok B.;
RT "Regulation of the aldehyde dehydrogenase gene (aldA) and its role in
RT the control of the co-inducer level necessary for induction of the
RT ethanol utilization pathway in Aspergillus nidulans.";
RL J. Biol. Chem. 276:6950-6959(2001).
DR EMBL; AF260125; AAK18074.1; -.
DR HSSP; P05091; ICW3.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
FT VARIANT 338 338 S->G.
SQ SEQUENCE 497 AA; 54158 MW; 54F425A3B3063875 CRC64;

Query Match 34.6%; Score 883; DB 3; Length 497;
Best Local Similarity 42.9%; Pred. No. 2.4e-60;
Matches 206; Conservative 78; Mismatches 174; Indels 22; Gaps 11;

QY 12 LFINNEFVSSKGSERLTLTPWDESTVATDVHVAADVDSAVAASVQAVKPKKKTG 71
DB 21 LFINNEFVKGVEGTFOVIPSNEK-VITSVHEATEKDDVDVAFAA-RAAFEGPWRQVTP 78
QY 72 ARAACMLKFDALAEKNAELARLESPTGRPVSMITHFDIPNMSVFYRAGWADKIG 131
DB 79 SERGILINKLADLMERDITDLAAIESLDNGKAFM-AKYDLANSIGLRYAGWADKING 137
QY 132 KTFEPDNGKPNW-RYEPWGVGAGTSNATFLVGVKTIAPALACGSFTFKASEKSPGV 190
DB 138 QITDITPETLTITRHEPVGCGQIIPNFPFLMWSMGIPVAAGNTVVLKTAETQIPUSA 197
QY 191 LGLAPFAEAGFPVGVQFTLG-ARVTGEALASHMDIAKISFTRSVGGGAVKQATLAKSN 249
DB 198 LYAAKLKEAGFPAGVINVISGFRAGAAISHMDIDKRVATGTLVGRITLIQAAKSN 257
QY 250 MKRVTLEIGEK-PTIVNEAPLERQ-SGESAKDFSKFGQIWWPPSCLLVQWGNLAEKF-- 305

```

```

DB 258 LKVTLELGGKSPNIVFDDADIDNAISWNFHGCCAGSRILVQEG-YYDKFVA 316
QY 306 ---HGVRHGSGFGCCQRLWGNLQONLEPKRTHGPFVDRKSOYDRVLGNIDVCKDT-AOLLTGVG 361
DB 317 REKERAOKNVG-----NPFQDTFQGPQVSOLOFDRIMEYINHKKAGATVATGSD 368
QY 362 RKDGKFAIEPTTFVNPCKSGKIWFEEIFGPVLSIKTFKTEETETIANDTDTYGLASVIY 421
DB 369 RHGNEGYFIQPTFTDVTSDMKIAQEEIFGPVVTIQKFDEAEAIKIGNSTIDYGLAAVH 428
QY 422 TKSINRGLRVSSALETGGVSNFPFIPETQTPFGMKQSGSGRELGEGLKAYLEPKTIN 481
DB 429 TKNVNTAIRVSNALKAGTVINNYNMISYQAPFGGKQSGLGRGSLGVALENTYQIKTVH 488

RESULT 5
Q9C2S1 PRELIMINARY; PRT; 519 AA.
AC Q9C2S1;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE 2700007F14RIK PROTEIN.
GN 2700007F14RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gofobori T., Bono H., Kasukawa T., Saito R.,
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gliss C., King B., Kochava H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirimi L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guncinich S., Hill D., Hofmann M., Hume D.A., Kaniya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK012213; BAB28101.1; -.
DR HSSP; P05091; ICW3.
DR MGD; MGI:1919785; 2700007F14RIK.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
SQ SEQUENCE 519 AA; 57552 MW; 69222D7409BFEF3 CRC64;

Query Match 34.6%; Score 882.5; DB 11; Length 519;
Best Local Similarity 41.6%; Pred. No. 2.8e-60;
Matches 207; Conservative 84; Mismatches 174; Indels 33; Gaps 12;

QY 4 SPEYKSELPFINNEFVSSKGSERLTLTPWDESTVATDVHVA--NAAVDSAVASVQAV 61
DB 33 NPEICYNKULFINNEWHDAVSKKTFPTVNTTGEVIG--HVAEGDRAVDVLAARAARAF 89
QY 62 KKG-PWKKTGAGRAACMLKFDALAEKNAELARLESPTGRPVSMITHFDIPNMSVYR 120
DB 90 RLCSPPWRMDASGRGLLNKLADLVERDRVYLALETLDNGKFPQESVYLDLDEVIKVR 149

```

QY 121 YYAGWADKAGKTFPEDNGKPNW-RYEPMGVACIASWNNATFLYGVGKAPALAAACSF 179
Db 150 YFAGWADKWHGKTFIPMOGEHFCFTRHEPVGVCQIIPNFPPLVMQGWKAPALATGNTVV 209
QY 180 FKASEKSPGLVGLAPLFAEAGFPVGVQVLTGARVT-GEALASHMDIAKISFTSRVGGG 238
Db 210 MKVAQTPLSALYLASLIKEAGFPVGVNITIGYGTAGANIAQHMDVDRKVAFTGSTEVG 269
QY 239 RAVKQATLKSNNKRVTLLEGEK-PTIVFNEAPLER---QSGESAKDFSKFGQIWWPPSCL 294
Db 270 HLIQKAAGESNLKRVTLLEGEKSPSIVLADADMEHAYDQCHEAL--FFNMGOCCEAGSRT 327
QY 295 LVOMG-----NLAKEFHGVHSGFCGRCORWLGONPLEPKRTHGPPVDKSOVDRLVGN 347
Db 328 FVEESIYREFLERIVEK-----AKORKVG-NPFDLTQGGQVQDKGEFERILGYI 376
QY 348 DVG-KDTAQLTGVGRKDGKGFATEPIFVNPFGKSKIWEIEIFGVLSTIKTFKTEEEAI 406
Db 377 RLGQKEGAKLLCGGERLGERGFFIKTIVFGDVQDGRIRAKKEEIFGVPFLFKFKTEEVI 436
QY 407 ETIANDTYGLASVITYKSLNGLRVSSALETGGVSINFPFIPETQTPFGGKMGKSGSREL 466
Db 437 ORANTRYGLAAAVETRLDKAIYFTQALQAGTWWNTYNIIVTCHTFPGGPNESGNREL 496
QY 467 GEGGLKAYLEPKTINIHW 484
Db 497 GEDGURATEVKTITIKV 514

RESULT 6
ID Q35945 PRELIMINARY; PRT; 501 AA.
AC Q35945;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE ALDEHYDE DEHYDROGENASE AHD-2-LIKE.
GN ALDH1A4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL; TISSUE=LIVER;
RA Hau L.C.; Chang W.C.; Yoshida A.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U96401; AAB64411.1; -.
DR HSSP; P51977; LBXS.
DR MGD; MGI:1347050; Aldh1a4.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
SQ SEQUENCE 501 AA; 54587 MW; 42DE97962799237B CRC64;

Query Match 34.5%; Score 878.5; DB 11; Length 501;
Best Local Similarity 41.9%; Pred. No. 5.5e-60;
Matches 204; Conservative 89; Mismatches 173; Indels 21; Gaps 11;

QY 10 SELFFINNEFVSKSERLTITNPDWDESTVATDVHANAADVDASAASVQAVKKG-PNKK 68
Db 21 TKIFINNEHDSVSSKFPVLPNATEVVIC-HVEEGDKADVDAVRAAQAQFQISPHWT 79
QY 69 FTGAORACMLKPADLAEKNAEKARLESUPTGRPSVMTHTFDIPNMVSVFFRYAGWADK 128
Db 80 MDASERGLLNKLLADLMERDRLLLATMESNAGKVFPAHAYLLIDVEISIKALQYFAGWADK 139
QY 129 IAGTTFPDNGKPNW-RYEPMGVACIASWNNATFLYGVGKAPALAAACSFIFKASEKSP 187
Db 140 IHGOTIPSDGNIYTYTRREPIGVCGIIPWNGPLIIFTWKLGPALSCGNVTVVVKAEQTP 199

QY 188 LGVLGLAPLFAEAGFPVGVQVLTGARVT-GEALASHMDIAKISFTSRVGGGKAVKQATL 246
Db 200 LTALHMSLIKEAGFPVGVNIVPGYGTAGGAISSHMDIDKVSFTGSTEVGKLIKEAG 259
QY 247 KSNMKRVTLLEGEK-PTIVFNEAPLERQSGESAKDFS-----KFGQIWWPPSCLLVQMG 299
Db 260 KSNLKRVTLEGEKSPCIVFADADL-----DSAVEFAHOGVFFHQGGICVAAASRLFVE-E 313
QY 300 NLAKEFHGVHSGFCGRCORWLGONPLEPKRTHGPPVDKSOYDRVLGNIDVG-KDTAQLLT 358
Db 314 SIYDEF--VRR-SVERAKKYILGNLSNGINQGPQIDKEQHNKILGLIESGKKEGAKLEC 370
QY 359 GVGKGDGKGFATEPIFVNPFGKSKIWEIEIFGVLSTIKTFKTEEEAIAIANDTYGLAS 418
Db 371 GGGWGNKGFVQTPVFSNVTDEMIRAKEIFGVPVQVQIMKFSKMDVIVKRAANTTYGLAA 430
QY 419 VIYTKSLNGLRVSSALETGGVSINFPFIPETQTPFGGKMGKSGSRELGEGLKAYLEPK 478
Db 431 GVTETDLDKAITVSSALQAGWVWVNCYLAVPVQCPFGGKMGSGNRELGEHGLYEYTELK 490
QY 479 TINIHN 485
Db 491 TVAQIS 497

RESULT 7
ID Q9P303 PRELIMINARY; PRT; 497 AA.
AC Q9P303;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ALDEHYDE DEHYDROGENASE.
GN ALDH.
OS Cladosporium fulvum (Fulvia fulva).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina;
OC Dothideomycetes et Chaetothyriomycetes Incertae sedis;
OC Mycosphaerellaceae; mitosporic Mycosphaerellaceae; Cladosporium.
OX NCBI_TaxID=5499;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CF4;
RX MEDLINE=98052127; PubMed-9390425;
RA Coleman M.; Henricot B.; Arnau J.; Oliver R.P.;
RT "Starvation-induced genes of the tomato pathogen Cladosporium fulvum are also induced during growth in planta."
RL Mol. Plant Microbe Interact. 10:1106-1109(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CF4;
RA Segers G.C.; Bradshaw N.; Archer D.; Blissett K.; Oliver R.P.;
RT "Alcohol oxidase is a novel pathogenicity factor for Cladosporium fulvum but aldehyde dehydrogenase is dispensable."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF275347; AAF82789.1; -.
DR HSSP; P05091; ICW3.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF001128; Cyt_P450.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN.1.
SQ SEQUENCE 497 AA; 53849 MW; 1B1EEB10FB85599 CRC64;

Query Match 34.2%; Score 872; DB 3; Length 497;
Best Local Similarity 42.1%; Pred. No. 1.7e-59;
Matches 207; Conservative 74; Mismatches 181; Indels 30; Gaps 12;

QY 7 EYKSELFINNEFVSKSERLTITNPDWDESTVATDVHANAADVDASAASVQAVKKG 66
Db 16 EQPTGLFINNEWKPVSQKFFDVINPTDESVC-QVAEADKDVDAVAAAKAF-EGPW 73

```
QY 67 KKFTGAQRAACMLKFAADLAENAKLARLESIPGTGRPVSMITHFDIPNMVSVFPRYAGWA 126
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 74 ROETPENRGKLLVLAELFERNLDLFASVKALDNGKAFNM-AKVDIGMCACGLRYYGWA 132
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 127 DKIAKTPPEDNGKFNRY----EPMGVCAGIASWNAFLYGVWKIAPALAAAGCSFIFKAS 183
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 133 DKIEGKV- -DTPDPTFNVIKKEPIGVCGQIIPNFPFLMAMKIGPAVATGNTVVIKTA 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 184 EKSPILGVLGLAPFAEAGFPVGVQFLTG-ARVTGEALASHMDIAKISFTRSVGGGRAVK 242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 EOTPLGAYVAANLKEAGFPVGVNVISGCKVAGAAASHMDIDKVAFTGSVTGTIL 250
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 243 QATLKSNNKRYTLELGER-PTIVNEAPLEHQ-SGESAKDFSKFGQIWPVPPSCLLVOMGN 300
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 251 KAAAGSNLKKVTLELGGKSPNIVFDDADIDNAISWNVFGIFFNHGRCRCCAGSIYVOEG- 309
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 301 LAEKPHGVHSGFGCCRWLGO-----NPLEPKRTHGPFVDKSOYDVLGNIDVGKDT 353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 310 IYDKF-----LQFKERAQKNVVGDPFAADTQGGVOVSQVQDRINGYIQAGRDA 359
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 354 -AOLLTGVGRKDGKFAIEPTIFVNPKPCKSKINFEIEFGPVLSIKTKTEEEAEIANDT 412
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 360 GATVEIGGNRKDGKGYFEPTIFANVTEDMKIMQEEIFGPPVCSISKFKTKEDAIKVGNA 419
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 413 TVGLASVIYTSLSNGLRVSALSTGGVSINFPETOTPEGCKMKGSGRELGEGLK 472
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 420 PYGLAAVAHTRNLNTAIEVSNALRAGTVWNTYNTLHHQLPFGFKESGIGRELGEDALA 479
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 473 AYLEPKTINIHV 484
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 480 NVTQTKVSI 491
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
Q9TXM0
ID Q9TXM0 PRELIMINARY; PRT; 537 AA.
AC Q9TXM0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 58.6 KDA PROTEIN.
GN K04F1.15.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
   investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018(1998).
RN [2];
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Tin-Wollam A., Hankhaus J., Wohldmann P., Osborn A.;
RT "The sequence of C. elegans Cosmid K04F1."
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [3];
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF106575; AAC78174.2; -.
DR HSSP: P05091; 1CW3.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldehyd; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_GLU; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Hypothetical protein.
```

```
SQ SEQUENCE 537 AA; 58601 MW; F88875C8B460003D CRC64;

Query Match
Best Local Similarity 41.7%; Pred. No. 5.4e-58;
Matches 203; Conservative 75; Mismatches 184; Indels 25; Gaps 11;

QY 11 ELFINNEFVSSKSGSRLLTITNPNWDESTVATDVHANAADVDSAYAAVSQAVKKG-PWKKF 69
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 58 OLFINNEFVDSKSGKTEEFVNPANGKLLA-KVAEGNRDDVDIAVEAAKAKFKIGSEWRM 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 70 TGAQRAACMLKPADLAENAKLARLESIPGTGRPVSMITHFDIPNMVSVFPRYAGWA 129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 DASHRGVLLNRDLMERDRVILASLESIDNGKPYKEAYNIDLPTISIKTERFYAGYADKN 176
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 130 AKCTPEDNGKPNW-RYEPMGVCAGIASWNAFLYGVWKIAPALAAAGCSFIFKASEKSP 188
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 177 HGKTIIPVGGDYFTYTRHEPVGCGQIIPNFPFLMAMKIGPAVATGNTVVIKTA 236
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 189 GVLGLAPFAEAGFPVGVQFLTGARV-TGEALASHMDIAKISFTRSVGGGRAVKQATLK 247
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 237 SALHVASLIKEAQFPVGVNIIIPGRGTDAEAIASHMDVDRKVAFTGSTEVEGKTIKAAAE 296
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 248 SNMKRVTLLELGEK-PTIVNEAPLE---ROSGESAKDFSKFGQIWPVPPSCLLVOMGNLAE 303
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 297 SNVKVTLLELGEKSPNIVFADADLEAVRQSHAL--FFNQGCCSAGSRTFVE-GKIYD 353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 304 KFHG-----VRHSGFGCCRWLGNLQNPBPKRTHGPFVDKSOYDVLGNIDVG-KDTAQLL 357
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 354 EFVAKAKELVEKTVIG-----DPFDENTQGPQIDESQVETIMKYIESCKKEGAQLV 405
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 358 TGVGRKDGKFAIEPTIFVNPKPCKSKINFEIEFGPVLSIKTKTEEEAEIANDT 417
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 406 TGGVRHGDQGYFVKPTIFANVNDQMKIAQEEIFGPMIVIRPDSMEELIEKANNTIYGLA 465
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 418 SVIYTKSLNRLGVSSALETGCVSINFPETOTPEGCKMKGSGRELGEGLKAYLEP 477
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 466 AGVVTNDLNLKALJOVANTIRAGSVVWVNCYDVPDPAAPFGGKQSGIGRELGEGLAAYTEV 525
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 478 KTIINH 484
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 526 KTVTKV 532
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
Q90XS8
ID Q90XS8 PRELIMINARY; PRT; 518 AA.
AC Q90XS8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE RETINALDEHYDE DEHYDROGENASE TYPE 2.
GN RALDH2.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1];
RP SEQUENCE FROM N.A.
RX MEDLINE=21543521; PubMed=11688558;
RA Begemann G., Schilling T.F., Rauch G.-J., Geisler R., Ingham P.W.;
RT "The zebrafish neckless mutation reveals a requirement for RALDH2 in
   mesodermal signals that pattern the hindbrain."
RL Development 128:3081-3094(2001).
DR EMBL: AF393837; AAL00899.1; -.
SQ SEQUENCE 518 AA; 56515 MW; 2C6C451617E8E086 CRC64;

Query Match
Best Local Similarity 41.8%; Pred. No. 1e-57;
Matches 205; Conservative 78; Mismatches 195; Indels 13; Gaps 11;
```


[illegible]

| | |
|--------|---|
| RESULT | 13 |
| Q918W8 | |
| ID | Q918W8 |
| AC | PRELIMINARY; |
| DC | PRT; 517 AA. |
| DT | Q918W8; |
| DT | 01-OCT-2000 (TrEMBLrel. 15, Created) |
| DT | 01-OCT-2000 (TrEMBLrel. 15, Last sequence update) |
| DT | 01-DEC-2000 (TrEMBLrel. 19, Last annotation update) |
| DE | CLASS I ALDEHYDE DEHYDROGENASE. |
| OS | Euphonia guttata (Zebra finch) (Taeniopygia guttata). |
| OS | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| OC | Archosauaria; Aves; Neognathae; Passeriformes; Estrilidae; |
| OC | Estrilidinae; Taeniopygia. |
| OX | NCBI_TaxID=59729; |
| RN | [1] |
| RP | SEQUENCE FROM N.A. |
| RX | MEDLINE=20439488; PubMed=10985355; |
| RX | Denisenko-Neubass N.I., Jarvis E., Scharff C., Nottebohm F., |
| RA | Mello C.V.; |
| RA | "Site-specific retinoic acid production in the brain of adult |
| RL | songbirds."; |
| RL | Neuron 27:359-370(2000). |
| RL | EMBL: AF162770; AAF80471.1; . |
| DR | HSSP; P51977; 1BX5. |
| DR | InterPro; IPR002086; Aldehyde_dehydr. |
| DR | pfam; PF00171; aldedh; 1. |
| DR | PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1. |
| DR | PROSITE; PS00687; ALDSHYDE_DEHYDR_GLU; 1. |
| SO | SEQUENCE 517 AA; 55595 MW; D842A2A1BF4DA109 CRC64; |

| | | | | |
|-----------------------|--------------|--------------------|-----------------|---------------------|
| Query Match | 32.4% | Score 825: | DB 13: | Length 517: |
| Best Local Similarity | 40.5%: | Prod. No. 8.3e-56: | | |
| Matches 195: | Conservative | 84: | Mismatches 190: | Indels 12: Gaps 11: |

| | | | | | | | |
|----|-----|--------------------|------------------|------------|-----------|------------|---------------------|
| QY | 10 | SELFNNFEVSKSGSERLT | LNPNDESTVATD | VHVANAADVD | SAVAASVOA | VKKGP-WKK | 68 |
| | | ::: ::: | ::: | ::: | ::: | ::: | ::: |
| Db | 38 | TKIFNNWQNS | ESGRIPV | NPNTG | QIC-DIO | ADKVD | TKVAARLAFSLGSWRR 96 |
| | | ::: ::: | ::: | ::: | ::: | ::: | ::: |
| QY | 69 | FTCAARAACMLK | FADEAKNAEKLARLES | LPTGR | VPYSMITH | FDIPNMVSVF | RYAGNADK 128 |
| | | ::: ::: | ::: | ::: | ::: | ::: | ::: |
| Db | 97 | MDASERGHLLDK | LADLVERORAILAT | WESLNSCK | PFLOAFYVD | LQGVIK | TLRYAGNADK 156 |
| | | ::: ::: | ::: | ::: | ::: | ::: | ::: |
| QY | 129 | IAGKTFPEDNCK | PNW-RYEP | MGVCAG | TASNNATFL | VYWGKIAPAL | AAGCSFTFFKASEKP 187 |
| | | ::: ::: | ::: | ::: | ::: | ::: | ::: |
| Db | 157 | IIGMTIPVDG | QFTFTRHEP | IGVCGQI | PNWFLLMFA | NKIAPALCCG | NTVYIKPAEQTP 216 |
| | | ::: ::: | ::: | ::: | ::: | ::: | ::: |

```
QY 188 LGVIGLAPLFAEAGFPVGVQVLTG-ARYTGEALASHMDIAKISFTRSVGGGRAYKQATL 246
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 217 LSALYMGALIKFAGFPVGVVNIPLFGCPVIVGAIAASHVGDIDKIAFTGTSTGVSKLQEAAG 276
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 247 KSNKRVLTLELGEK-PTIVFNEAPLERQSGESAKD--FSKFGQIWWPPSCLLVQWGNLAEK 304
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 277 RSNLKRVTLELGGKSPNIIADADLDYAVQAHQGVFFNQGCCAGTSRIYVE-ESIYEE 335
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 305 FHGVHSGFGCCQWNLGQNPLEKPRTHGPPVKSQSDRVLGNDVG-KDTAQLLGVGRK 363
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 336 F--VRK-SVRAKRKIVGSPFDTTEQGPQIDKKQVKNKILEIQSGITBGALECGGKGL 392
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 364 GDKFAIEPTIFVNPKPGSKIMFEEIFGPVLSIKTFKTEETEEATEIANDTTYGLASVIYTK 423
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 393 GRKGFEIEPTVFSNVTDMDRIAKKEIFGPVQELIRFETWDEVIERANNSDFGLVAFTN 452
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 424 SLNRGLVSSALETGGVSNINFPPIPETQPPFGMKQSGSGRELGRGLKAYLEPKTINI 483
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 453 DINKALTSSAMQAGTWINCYNALNAQSPFGGSK-SGNGREMGEGCLREYSEVATVITK 511
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 484 V 484
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 512 I 512
```

```
RESULT 14
QYGY2 ID QYGY2 PRELIMINARY: PRT; 502 AA.
AC QYGY2;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE ALDEHYDE DEHYDROGENASE (EC 1.2.1.3).
GN ALDHL.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99191703; PubMed=10091603;
RA Ang H. L... Duester G.;
RT "Stimulation of premature retinoic acid synthesis in Xenopus embryos
following premature expression of aldehyde dehydrogenase ALDHL.";
RL Eur. J. Biochem. 260:227-234(1999).
DR EMBL; AF061833; AAC69552.1; -.
DR HSSP; P51977; 1BX5.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase.
SQ SEQUENCE 502 AA; 55205 MW; 5466BCDB85CAA7D6 CRC64;
```

Query Match 32.2%; Score 822; DB 13; Length 502;
Best Local Similarity 40.5%; Pred. No. 1.4e-55;
Matches 199; Conservative 90; Mismatches 188; Indels 14; Gaps 12;

```
QY 3 LSPDEYK-SELFTNNEFVSKSGSERLTLPNPDVESTVATDVHANAADVDSAVAASVOAV 61
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 14 LSNLEIKYTKIFITNNMHNHNSISGKFPVYPATGEKVC-EVEEGDKEDVNAKVAAREAF 72
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 62 KKG-PWKKFTGAORACMLKFAADLAENAEKLARLESPTGRPVSMITHFDIPNMVSVFR 120
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 73 QIGSPWRLLDASGRMLNKLADLVERDRLLISTMESIDGKPYASYFGDIPGAKSLR 132
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 121 YIAGWADKIAGKTFPEDNGKPNW-RYEPMGVCAGTASNNATFLYGVWIAAPALAGCSFI 179
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 133 YCAGWADKVOGTIPMDGYFTFTRHEPVGVCGQIIPNWFPLVMPFAWKIAPALCCGNTV 192
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
```

```
QY 180 FKASEKSPGLVGLAPLFAEAGFPVGVQVLTGARVT-GEALASHMDIAKISFTRSVGGG 238
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 193 IKPAEQITPTALYMGSLIKFAGFPVGVVNIPLFGCPVIVGAIAASHVGDIDKIAFTGTSTGV 252
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 239 RAYKQATLSNKRVTLELGEK-PTIVFNEAPLERQSGESAKD--FSKFGQIWWPPSCLL 295
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 253 KLKIEAAGKSNLKRVTLELGGKSPNIIADADLD-LAVEHAHNGLPFHQGCCIAGRIF 311
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 296 VQNGNLAERPHGVHSGFGCCQWNLGQNPLEKPRTHGPPVKSQSDRVLGNDVG-KDTA 354
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 312 VE-EPTYDEF--VRKSVRAKRKRVLG-DPFAPCVQNPQIDKEQYDKCLELJESGKKEGA 367
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 355 QLLTGVGRKGDGKFAIEPTIFVNPKPGSKIMFEEIFGPVLSIKTFKTEETEEATEIANDTTY 414
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 368 KLOGGSGANGKEGFIYSPVFSVDKDDRIAKKEIFGPVQVQILKPKTIDEVTKRANNTKY 427
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 415 GLASVIYTSLSNRLGRVSSALETGGVSNINFPPIPETQPPFGMKQSGSGRELGRGLKAY 474
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 428 GLAAGVFTKMDKAILMTALQATVYINCYNSAMSPQSPFGGKMGNGREMGEGYGLHEY 487
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 475 LEPTKINIHN 485
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 488 TEVKTVMKIS 498
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
```

```
RESULT 15
QY4293 ID QY4293 PRELIMINARY: PRT; 503 AA.
AC QY4293;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PROBABLE ALDEHYDE DEHYDROGENASE (EC 1.2.1.3) (ALDDH).
GN SPAC9E9.09C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RA McDougall R., Barrell B.G., Rajandream M.A., Wood V.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -I- CATALYTIC ACTIVITY: ALDEHYDE + NAD(+) + H(2)O -> ACID + NADH.
CC -I- PATHWAY: SECOND STEP IN ETHANOL UTILIZATION.
CC -I- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DR EMBL; Z99262; CAB16407.1; -.
DR HSSP; P05091; 1CW3.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD.
FT NP_BIND 247 252 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 270 270 BY SIMILARITY.
FT ACT_SITE 304 304 BY SIMILARITY.
SQ SEQUENCE 503 AA; 54768 MW; A7787A2181FB9CD5 CRC64;
```

Query Match 32.2%; Score 820; DB 3; Length 503;
Best Local Similarity 39.5%; Pred. No. 2e-55;
Matches 195; Conservative 85; Mismatches 178; Indels 36; Gaps 12;

```
QY 12 LFTNNEFVSKSGSERLTLPNPDVESTVATDVHANAADVDSAVAASVOAVK-KGPKKFT 70
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 25 LFINNHQVDSVHGGRVKVYSPSTEKLIC-EVADAEEDVDIAVKVAAAFOTDAPWRKFS 83
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 71 GAORACMLKFAADLAENAEKLARLESPTGRPVSMITHFDIPNMVSVFRYAGWADKIA 130
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 84 SAORGLSLRLADEQNLEYLASIETLDNGKSIIT-LARGDVAADCFRYGGWADKDY 142
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 131 GKTFFPDNGKPNW-RYEPMGVCAGTASNNATFLYGVWIAAPALAGCSFIFKASEKSLG 189
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
```

Db 143 GOTTETDIKRYATRHPIGVCGQIIPWFFFLCMKRIAPVACGNTIILKTAEITPLS 202
QY 190 VLGLAPLFAEAGFPFGVYQFLTG-ARVTGEALASHMDIAKISFTSRVGGGAVKQATLKS 248
Db 203 ALCLTKFYPCGFPFGVINVLSDGRRGCGNAISSHMDIDKVAFTGTGVRMYMRAAAS 262
QY 249 NMKRVTLLELGEK-PTIVFNEAPLERQSGESAKDFSKF-----CQIWVPPSCLLVQWGNL 301
Db 263 NLKKVTLLELGGKSPNIVENDADL-----DSAAVNTNYGIFYNSSGVCCAGSRVYVQ-EDV 316
QY 302 AEKPHGVHRHSGFCGCRWLGO-----NPLEPKRTHGPFVDSQYDRVLGNIDVG-KDT 353
Db 317 YDEP-----IKMKAKAKTLKVGDPFAEDTFQGAQVSKQOYERIVSYIESIAHG 366
QY 354 AOLLTGVRKGDGKFAIEPTIFVNPKPQSKIWFEIIPGVLUSIKTKFTEEAIEIANOTT 413
Db 367 AKLEIGGKRHNGLGYFEPTILSNVTEDMAVGKEEIFGPVLAVTKFKTIEEAIIRGNNST 426
QY 414 YGLASVIYTKSLNRGLRVSSALETGCVSINFPPIPETOTPEGGMKQSGSGRELGECLKA 473
Db 427 YGLAAGVHTNITNAIKVSNALFAGTVVWVNCYNLLHHQIPFGYKESGIGRELGSYGLTN 486
QY 474 YLEPKTINIHVNIE 487
Db 487 YTOTKAVHNLGMD 500

Search completed: April 27, 2002, 07:57:44
Job time: 700 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:59:37 ; Search time 35.22 Seconds
(without alignments)
535.390 Million cell updates/sec

Title: US-09-882-694A-5

Perfect score: 2549

Sequence: 1 MVLSPEYKSEKLFNNEFVS.....EEGLKAYLEPKTINHVNI 487

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Query % | Length | DB ID | Description |
|------------|-------|-------|---------|--------|------------|---------------------|
| 1 | 882.5 | 34.6 | 517 | 1 | DHA5_HUMAN | P30837 homo sapien |
| 2 | 866 | 34.0 | 496 | 1 | DHAL_CLAHE | P40108 cladosporiu |
| 3 | 864 | 33.9 | 497 | 1 | DHAL_EMENI | P08157 emericella |
| 4 | 855.5 | 33.6 | 500 | 1 | DHAL_HORSE | P15437 equus caball |
| 5 | 855.5 | 33.6 | 500 | 1 | DHAL_MOUSE | P24549 mus musculus |
| 6 | 855.5 | 33.6 | 500 | 1 | DHAL_MOUSE | P13601 rattus norv |
| 7 | 848.5 | 33.3 | 500 | 1 | DHAL_RAT | P51647 rattus norv |
| 8 | 846.5 | 33.2 | 499 | 1 | DHAL_HUMAN | P04788 homo sapien |
| 9 | 843.5 | 33.1 | 500 | 1 | DHAL_HUMAN | P00352 homo sapien |
| 10 | 842 | 33.0 | 497 | 1 | DHAL_ASPEC | P41751 aspergillus |
| 11 | 840.5 | 33.0 | 499 | 1 | DHAL_CHICK | O93344 gallus gall |
| 12 | 840.5 | 33.0 | 517 | 1 | DHAL_HUMAN | P05091 homo sapien |
| 13 | 838.5 | 32.9 | 501 | 1 | DHAL_MACPR | Q29490 macroscelid |
| 14 | 838.5 | 32.9 | 509 | 1 | DHAL_CHICK | P27463 gallus gall |
| 15 | 836.5 | 32.8 | 499 | 1 | DHAL_MOUSE | Q62148 mus musculus |
| 16 | 835.5 | 32.8 | 499 | 1 | DHAL_MOUSE | Q63639 rattus norv |
| 17 | 834.5 | 32.7 | 519 | 1 | DHAL_MOUSE | P47738 mus musculus |
| 18 | 833.5 | 32.7 | 519 | 1 | DHAL_RAT | P11884 rattus norv |
| 19 | 832.5 | 32.7 | 501 | 1 | DHAL_ELEED | Q28399 elephantulu |
| 20 | 830.5 | 32.6 | 500 | 1 | DHAL_SHEEP | P51977 ovis aries |
| 21 | 829.5 | 32.5 | 520 | 1 | DHAL_BOVIN | P20000 bos taurus |
| 22 | 826.5 | 32.4 | 500 | 1 | DHAL_BOVIN | P48644 bos taurus |
| 23 | 826.5 | 32.4 | 500 | 1 | DHAL_MESAU | P81178 mesocricetu |
| 24 | 824.5 | 32.3 | 512 | 1 | DHAL_HUMAN | P47895 homo sapien |
| 25 | 815.5 | 32.0 | 495 | 1 | DHAL_ALTAL | P42041 alternaria |
| 26 | 810.5 | 31.8 | 497 | 1 | DHAL_ENCBU | Q27640 enchytraeus |
| 27 | 808.5 | 31.7 | 500 | 1 | DHAL_HORSE | P12762 equus caball |
| 28 | 775 | 30.4 | 500 | 1 | DHAL_AGABI | O74187 agarcus bi |
| 29 | 772.5 | 30.3 | 495 | 1 | CROM_OCTDO | P30841 octopus dof |
| 30 | 768 | 30.1 | 498 | 1 | CROM_OMMSL | P30842 omastreph |
| 31 | 727 | 28.5 | 494 | 1 | DHAL_LEITA | Q25417 leishmania |
| 32 | 724.5 | 28.4 | 495 | 1 | DHAL_ECOLI | P23883 escherichia |
| 33 | 723.5 | 28.4 | 490 | 1 | DHAL_BACSU | P71016 bacillus su |

| | | | | | | |
|----|-------|------|-----|---|------------|--------------------|
| 34 | 715 | 28.1 | 519 | 1 | DHA5_YEAST | P40047 saccharomyc |
| 35 | 707.5 | 27.8 | 501 | 1 | DHAB_AMAHP | O04895 amaranthus |
| 36 | 707.5 | 27.8 | 506 | 1 | DHA2_YEAST | P47771 saccharomyc |
| 37 | 707 | 27.7 | 902 | 1 | FTDH_HUMAN | O75891 homo sapien |
| 38 | 703 | 27.6 | 902 | 1 | FTDH_RAT | P28037 rattus norv |
| 39 | 702 | 27.5 | 497 | 1 | DHAB_SPIOL | P17202 spinacia ol |
| 40 | 699.5 | 27.4 | 499 | 1 | FEAB_ECOLI | P80668 escherichia |
| 41 | 696.5 | 27.3 | 500 | 1 | DHA6_YEAST | P54115 saccharomyc |
| 42 | 693 | 27.2 | 519 | 1 | DHA4_YEAST | P46367 beta vulgar |
| 43 | 692.5 | 27.2 | 500 | 1 | DHAB_BETVU | P28237 beta vulgar |
| 44 | 690.5 | 27.1 | 506 | 1 | DHA3_YEAST | P54114 saccharomyc |
| 45 | 675.5 | 26.5 | 502 | 1 | DHAB_ATRHO | P42757 atriplex ho |

ALIGNMENTS

| | | | | | | |
|------------|--|-----------|------|---------|--|--|
| RESULT 1 | | | | | | |
| DHA5_HUMAN | | | | | | |
| ID | DHA5_HUMAN | STANDARD: | PRT: | 517 AA. | | |
| AC | P30837: | | | | | |
| DT | 01-JUL-1993 (Rel. 26, Created) | | | | | |
| DT | 01-JUL-1993 (Rel. 26, Last sequence update) | | | | | |
| DT | 01-MAR-2002 (Rel. 41, Last annotation update) | | | | | |
| DE | Aldehyde dehydrogenase X, mitochondrial precursor (EC 1.2.1.3) (ALDH class 2). | | | | | |
| GN | ALDH1B1 OR ALDH5 OR ALDHX. | | | | | |
| OS | Homo sapiens (Human). | | | | | |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | | | | |
| OC | Mammalia; Euthera; Primates; Catarrhini; Homidae; Homo. | | | | | |
| OX | NCBI_TaxID=9606; | | | | | |
| RN | [1]_TaxID=9606; | | | | | |
| RP | SEQUENCE FROM N.A. | | | | | |
| RC | TISSUE=Testis; | | | | | |
| RX | MEDLINE=91286241; PubMed=2061311; | | | | | |
| RA | Hsu L.C., Chang W.-C.; | | | | | |
| RT | "Cloning and characterization of a new functional human aldehyde dehydrogenase gene."; | | | | | |
| RL | J. Biol. Chem. 266:12257-12265(1991). | | | | | |
| RN | [2] | | | | | |
| RP | VARIANTS VAL-86 AND ARG-107 | | | | | |
| RX | MEDLINE=94063858; PubMed=8244338; | | | | | |
| RA | Sherman D., Dave V., Hsu L.C., Peters T.J., Yoshida A.; | | | | | |
| RT | "Diverse polymorphism within a short coding region of the human aldehyde dehydrogenase-5 (ALDH5) gene."; | | | | | |
| RL | Hum. Genet. 92:477-480(1993). | | | | | |
| CC | -1- FUNCTION: ALDH5 PLAY A MAJOR ROLE IN THE DETOXIFICATION OF ALCOHOL-DERIVED ACETALDEHYDE. THEY ARE INVOLVED IN THE METABOLISM OF CORTICOSTEROIDS, BIOGENIC AMINES, NEUROTRANSMITTERS, AND LIPID PEROXIDATION. | | | | | |
| CC | -1- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O -> an acid + NADH. | | | | | |
| CC | -1- PATHWAY: SECOND STEP IN ETHANOL UTILIZATION. | | | | | |
| CC | -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY). | | | | | |
| CC | -1- SUBCELLULAR LOCATION: Mitochondrial matrix. | | | | | |
| CC | -1- TISSUE SPECIFICITY: LIVER, TESTIS AND IN A LESSER EXTENT IN BRAIN. | | | | | |
| CC | -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY. | | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). | | | | | |
| CC | EMBL: M63967; AAA96830.1; . | | | | | |
| CC | HSSP: P20000; 1AG8. | | | | | |
| CC | MIM: 100670; . | | | | | |
| CC | InterPro: IPR002086; Aldehyde_dehydr. | | | | | |
| CC | Prm: PF00171; atedh; 1. | | | | | |
| CC | PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1. | | | | | |
| CC | PROSITE: PS00687; ALDEHYDE_DEHYDR_GLD; 1. | | | | | |
| CC | Oxidoreductase; NAD; Mitochondrion; Transit peptide; Polymorphism. | | | | | |

[illegible]

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; LA2009; AAA96657.1; --
 DR EMBL; AF001896; AAC53304.1; --
 DR EMBL; AF001898; AAC53306.1; --
 DR EMBL; AF001897; AAC53305.1; --
 DR EMBL; U79118; AAB63423.1; --
 DR HSP; P51977; 1BX5.
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldedh; 1.
 DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
 DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 KW Oxidoreductase; NAD; Acetylation.
 FT INIT_MET 0 0
 FT MOD_RES 1 1 ACETYLATION (POTENTIAL).
 FT NP_BIND 245 250 NAD (ADP PART) (BY SIMILARITY).
 FT ACT_SITE 268 268 BY SIMILARITY.
 FT ACT_SITE 302 302 BY SIMILARITY.
 FT CONFLICT 99 99 R -> C (IN REF. 1).
 FT CONFLICT 105 105 I -> M (IN REF. 4).
 FT CONFLICT 169 169 N -> E (IN REF. 1).
 FT CONFLICT 169 169 N -> E (IN REF. 1).
 SQ SEQUENCE 500 AA; 54327 MW; 93614C21A94B430D CRC64;

Query Match 33.3%; Score 848.5; DB 1; Length 500;
 Best Local Similarity 40.68; Pred. No. 7.2e-59;
 Matches 196; Conservative 86; Mismatches 188; Indels 13; Gaps 8;
 QY 10 SELFNNFEVSKGSRLLTNPWDESTVATDVHANAADVDVAASVQAVKKG-PWK 68
 DB 20 TKIFINNEWHSDVSGKFPVLPNATEVIC-HVEEGDKADVKAKAARQAFQIGSPWRT 78
 QY 69 FTGAORAAACMLKADLAEKNAEKARLESPTGRPVSMITHFDIPNMVSVFRYAGWADK 128
 DB 79 MDASERGRLLNKLADLMERDRLLATTEALNGGKVFANAYLSDLGSSIKALKYAGWADK 138
 QY 129 TAGKTFPEDNCKPNW-RYEPMGVCAGIASWNATFLYGVKTIAPALAAAGSFIFKASEKSP 187
 DB 139 IHGQTIPSDGDIFFTRREPIGCGQIIPWNFPFLPWIKIGPALSCTGNTVYVKKAEQTP 198
 QY 188 LGVLGLAPLFAEAGFPVGVQFLTGARVT-GEALASHMDIAKISFTSRVGGGRVYKQATL 246
 DB 199 LTALHMASLIKEAGFPVGVNIPVPGYPTAGAAISSHMDVDKVAFTGSTQVGLIKEAAG 258
 QY 247 KSNMKRVTLGEGK-PTIVNEAPLERQSGESAKDFSKFGQIWPVPSCLLVQWNLAEK- 304
 DB 259 KSNLRVTLGEGKSPCTVADADL-----DIAVEFAHGVFHYHQCCVAASTRFVEES 313
 QY 305 -FHGVHSGFGCGQRWLNQNPLEKRTFGPVDKSDYRVLGNIDVG-KDTAQLLTGVGR 362
 DB 314 VYDFEVKSVSRERAKKYVGLNPLTQGINQGPQIDKEQDRKILDLIESKKKEGAKLCCGG 373
 QY 363 KDGKFAETPTIFVNPFGKINFEIFGPVLSIKTFKEEAEIANDTTTYGLASVYIT 422
 DB 374 WGNKGFVQPTVFSNVTDMRIAKEEIFGPVQOIMKFSIDDDVIRANNNTTGLAAGVFT 433
 QY 423 KSNLRVLSVALETGGVSNFFPIPTQTPFGCMQKSSGRELGEGLKAYLEKPTNI 482
 DB 434 KDLDRATVSSALQAGVWVWNCYNILLSAQCPFGGFRMSNGRELGHGLYEITELKTVM 493
 QY 483 HVN 485
 DB 494 KIS 496
 RESULT 8

DHA2_HUMAN STANDARD; PRT; 499 AA.
 ID DHA2_HUMAN
 AC O94788; 1999 (Rel. 38, Created)
 DT 19-JUL-1999 (Rel. 38, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Aldehyde dehydrogenase 1A2 (EC 1.2.1.3) (Retinaldehyde-specific
 DE dehydrogenase type 2) (RALDH(II)) (RALDH-2).
 GN ALDH1A2 OR RALDH2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99038200; PubMed=9819382;
 RA Ono Y., Fukuhara N., Yoshie O.;
 *TALI and LIM-only proteins synergistically induce retinaldehyde
 RT dehydrogenase 2 expression in T-cell acute lymphoblastic leukemia by
 RT acting as cofactors for GATA3 *;
 RL Mol. Cell. Biol. 18:6939-6950(1998).
 CC -| FUNCTION: RECOGNIZES AS SUBSTRATES FREE RETINAL AND CELLULAR
 CC RETINOL-BINDING PROTEIN-BOUND RETINAL. DOES METABOLIZE OCTANAL AND
 CC DECANAL BUT DOES NOT METABOLIZE CITRAL, BENZALDEHYDE, ACETALDEHYDE
 CC AND PROPANAL EFFICIENTLY (BY SIMILARITY).
 CC -| CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
 CC -| PATHWAY: RETINOIC ACID BIOGENESIS.
 CC -| SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
 CC -| SUBCELLULAR LOCATION: Cytoplasmic.
 CC -| SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB015226; BAA34785.1; ALT_INIT.
 DR HSP; P51977; 1BX5.
 DR MIM; 603687; --
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldedh; 1.
 DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
 DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 KW Oxidoreductase; NAD.
 FT NP_BIND 244 249 NAD (ADP PART) (BY SIMILARITY).
 FT ACT_SITE 267 267 POTENTIAL.
 FT ACT_SITE 301 301 POTENTIAL.
 SQ SEQUENCE 499 AA; 54780 MW; FE0B53A47644246B CRC64;
 Query Match 33.2%; Score 846.5; DB 1; Length 499;
 Best Local Similarity 41.0%; Pred. No. 1e-58;
 Matches 197; Conservative 84; Mismatches 189; Indels 11; Gaps 10;
 QY 10 SELFNNFEVSKGSRLLTNPWDESTVATDVHANAADVDVAASVQAVKKG-PWK 68
 DB 19 TKIFINNEWHSDVSGKFPVLPNATEVIC-HVEEGDKADVKAKAARQAFQIGSPWRT 77
 QY 69 FTGAORAAACMLKADLAEKNAEKARLESPTGRPVSMITHFDIPNMVSVFRYAGWADK 128
 DB 78 MDASERGRLLNKLADLMERDRLLATTEALNGGKVFANAYLSDLGSSIKALKYAGWADK 137
 QY 129 TAGKTFPEDNCKPNW-RYEPMGVCAGIASWNATFLYGVKTIAPALAAAGSFIFKASEKSP 187
 DB 138 IHGQTIPSDGDIFFTRREPIGCGQIIPWNFPFLPWIKIGPALSCTGNTVYVKKAEQTP 197
 QY 188 LGVLGLAPLFAEAGFPVGVQFLTGARVT-GEALASHMDIAKISFTSRVGGGRVYKQATL 246
 DB 198 LSALYMGALIKEAGFPVGVNIPVPGYPTAGAAIASHIGIDKIAFTGSTVEGKLIQEAAG 257

QY 247 KSNKRVTELEGEK-PTIVFNEAPLEROGESAKD-PSKFGQIHWPPSCLLVONGNLAEK 304
 Db 258 RSNLKVTELEGGKSNIIIFADADLDYAVEAHOGVFFNQCCCTAGSRIFVE-ESIYEE 316
 QY 305 FHGVRHSGSGGQWLGQNPLEKRPKTRHGFVVDKSDYDRVLGNIDVG-KDTAQLITGVGRK 363
 Db 317 F--VRSVERAKRRIVG-SPPDPTTEOGFOIDKKQYNKILEIQSGVAEGAKLECGGKGL 373
 QY 364 GDGFGFAIEFTIVNPKPGSKINFEELGFLVLSIKTEKTEBEAIEIANDTYYGLASVLYTK 423
 Db 374 GRGFFIEFTVPSNVYDDMKIAKEIFGVBOELRKFMDVIEIRANNSDFGLVAAVFTN 433
 QY 424 SLNRGLRVSALETGCVSINFFPETQTPFGMKQSGSGRELGEGLKAYLEKPTINIH 483
 Db 434 DINKALTVSSAQAQTVWNCYNALNAQSPFGGFKMSNGRMGEMFGLREYSEVKTIVTK 493
 QY 484 V 484
 Db 494 I 494

RESULT 9
 DHAI_HUMAN STANDARD; PRT; 500 AA.
 AC P00352; O00768;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Aldehyde dehydrogenase 1A1 (EC 1.2.1.3) (Aldehyde dehydrogenase,
 cytosolic) (ALDH class 1) (ALDH1) (ALDH-E1).
 GN ALDH1A1 OR ALDH1 OR ALDC.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-90077427; PubMed-2591967;
 RX Hsu L.C., Chang W.-C., Yoshida A.;
 RT "Genomic structure of the human cytosolic aldehyde dehydrogenase
 gene";
 RL Genomics 5:857-865(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE-94027752; PubMed-82144422;
 RA Zheng C.F., Wang T.T., Weiner H.;
 RT "Cloning and expression of the full-length cDNAs encoding human liver
 class 1 and class 2 aldehyde dehydrogenase";
 RL Alcohol. Clin. Exp. Res. 17:828-831(1993).
 RN [3]
 RP SEQUENCE OF 161-500 FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE-85216574; PubMed-2987944;
 RA Hsu L.C., Tani K., Fujiyoshi T., Kurachi K., Yoshida A.;
 RT "Cloning of cDNAs for human aldehyde dehydrogenases 1 and 2";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:3771-3775(1985).
 RN [4]
 RP SEQUENCE OF 161-500 FROM N.A.
 RX MEDLINE-85252089; PubMed-4015823;
 RA Yoshida A., Ikawa M., Hsu L.C., Tani K.;
 RT "Molecular abnormality and cDNA cloning of human aldehyde
 dehydrogenases";
 RL Alcohol 2:103-106(1985).
 RN [5]
 RP SEQUENCE.
 RC TISSUE=Liver;
 RX MEDLINE-84208019; PubMed-6723659;
 RA Hempel J., von Bahr-Lindstrom H., Joernvall H.;
 RT "Aldehyde dehydrogenase from human liver. Primary structure of the
 cytoplasmic isoenzyme";
 RL Eur. J. Biochem. 141:21-35(1984).
 RN [6]

RP ACTIVE SITE STUDIES, AND SEQUENCE OF 265-272.
 RX MEDLINE-88050861; PubMed-3676276;
 RA Abriola D.P., Fields R., Stein S., Mackerell A.D. Jr., Pietruszko R.;
 RT "Active site of human liver aldehyde dehydrogenase.";
 RL Biochemistry 26:5679-5684(1987).
 RN [7]
 RP PARTIAL SEQUENCE.
 RC TISSUE=Erythrocyte;
 RX MEDLINE-89377753; PubMed-2776714;
 RA Agarwal D.P., Cohn P., Goedde H.W., Hempel J.;
 RT "Aldehyde dehydrogenase from human erythrocytes: structural
 relationship to the liver cytosolic isozyme.";
 RL Enzyme 42:47-52(1989).
 RN [8]
 RP SEQUENCE OF 1-5 FROM N.A.
 RX MEDLINE-93263033; PubMed-8493914;
 RA Yoshida A., Hsu L.C., Yanagawa Y.;
 RT "Biological role of human cytosolic aldehyde dehydrogenase 1:
 hormonal response, retinal oxidation and implication in testicular
 feminization.";
 RL Adv. Exp. Med. Biol. 328:37-44(1993).
 CC -!- FUNCTION: BINDS FREE RETINAL AND CELLULAR RETINOL-BINDING PROTEIN-
 BOUND RETINAL. CAN CONVERT/OXIDIZES RETINALDEHYDE TO RETINOIC ACID
 (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O -> an acid + NADH.
 CC -!- PATHWAY: SECOND STEP IN ETHANOL UTILIZATION.
 CC -!- SUBUNIT: HOMOTETRAMER.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 DR EMBL; M31994; AAA51692.1;
 DR EMBL; M31982; AAA51692.1; JOINED.
 DR EMBL; M31983; AAA51692.1; JOINED.
 DR EMBL; M31984; AAA51692.1; JOINED.
 DR EMBL; M31985; AAA51692.1; JOINED.
 DR EMBL; M31986; AAA51692.1; JOINED.
 DR EMBL; M31987; AAA51692.1; JOINED.
 DR EMBL; M31988; AAA51692.1; JOINED.
 DR EMBL; M31989; AAA51692.1; JOINED.
 DR EMBL; M31990; AAA51692.1; JOINED.
 DR EMBL; M31991; AAA51692.1; JOINED.
 DR EMBL; M31992; AAA51692.1; JOINED.
 DR EMBL; AF003341; AAC51652.1;
 DR EMBL; K03000; AAA51695.1;
 DR EMBL; M26761; AAA35518.1;
 DR EMBL; S61435; AAD13925.1;
 DR PIR; A33371; DEH0EL.
 DR HSSP; P51977; IBXS.
 DR SWISS-2DPAGE; P00352; HUMAN.
 DR MIM; 100640;
 DR InterPro; IPR02086; Aldehyde_dehydr.
 DR Pfam; PF00171; aldedh; 1.
 DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
 DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 KW Oxidoreductase; NAD; Acetylation.
 FT INIT_MET 0 0
 FT MOD_RES 245 1 ACETYLATION.
 FT NP_BIND 268 250 NAD (ADP PART) (BY SIMILARITY).
 FT ACT_SITE 268 268 POTENTIAL.
 FT ACT_SITE 302 302 POTENTIAL.
 FT CONFLICT 120 120 N -> S (IN REF. 2).
 FT CONFLICT 161 161 V -> I (IN REF. 3 AND 4).
 SO SEQUENCE 500 AA; 54730 MW; 826464DC64C496C7 CRC64;

```
Query Match 33.1%; Score 843.5; DB 1; Length 500;
Best Local Similarity 39.8%; Pred. No. 1.8e-58;
Matches 192; Conservative 90; Mismatches 188; Indels 13; Gaps 8;

QY 10 SELFNNFVSSKGERLTLNPNWDESTVATDVHVAANAADVSAAVQAVKKG-PWK 68
DB 20 TKTFINNEHDSVSKKFPFVNPATEELC-QVEEGDKEDVAKVAAKQAFQIGSPWRT 78
QY 69 FTGAQRAACMLKFAADLAENAEKLAERLESPTGRPVSMITHFDIPNNVSVRYAGWADK 128
DB 79 MDASERGLLYKLADIERDRLLATNEMNGKLYSNAYLNDLACIKTLRVCAGWADK 138
QY 129 IAGKTFEPD-NCKPNWRPEPCVAGTASWATFLYGVNKAIPALAGCSFTFKASEKSP 187
DB 139 IGRITPIDGNFTTTRHEPIGVCGQIIPNFWPLMIWKIGPALSCGNTVVVVKPAEQTP 198
QY 188 LGVLGLAPLAFARAGPPGVQFLTGART-GEALASHMDIAKTSFTRSVGGGKRAVQAATL 246
DB 199 LTALHVASLIKRAAGPPGVWNVPGYGTAGAAISSHMDTKVAFTGSTEVGKLKEAAG 258
QY 247 KSNMKRVTLLEGEK-PTIVNEAPLERQSGESAKDFSKFGQIWWPPSCILLVQGNLARK- 304
DB 259 KSNMKRVTLLEGEKSPCLVADADL-----DNAVEFAHGVFPYHOGCCIAASRFPVES 313
QY 305 -FHGVHSGFGCCQRLQCNPLEPKRTHGPFVDSQTDRLVGNIDVG-KDTAQLLTGVR 362
DB 314 IYDFVRRSVRAKKYIILGNPLTPGVTOGQIDKEQYDKILDIESKKEGAKLEGCGGP 373
QY 363 KGDGFAIRPTTFVNPKPCKSWPEEIPFPVLSIKTKTEFEAEIANDTTTGLASVIYT 422
DB 374 WGNKGVFGVQPTVFSNVTDEMIAKEEIPFPVQIMKFKSLDDVIKRNNTFYGLSAGVT 433
QY 423 KSLNGLRVSSALGSGVSNFPLPETOTPFCKMGKSGSGRELGEGLKAYLEPKTINI 482
DB 434 KQIDKATISSALQASTVWVCYGVWSNOCFFGFKMGNGRELGEYGHETEVKTVTV 493
QY 483 HVN 485
DB 494 KTS 496

RESULT 10
DHALASPNQ STANDARD; PRT; 497 AA.
AC P41751.
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE Aldehyde dehydrogenase (EC 1.2.1.3) (ALDDH).
GN ALDA.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillius.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90108706; PubMed=2606357;
RA O'Connell M.J., Kelly J.M.;
RT "Physical characterization of the aldehyde-dehydrogenase-encoding
  gene of Aspergillus niger.";
RL Gene 84:173-180(1989).
CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O -> an acid + NADH.
CC -!- PATHWAY: SECOND STEP IN ETHANOL UTILIZATION.
CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
```

```
DR EMBL: M32351; AAA87596.1; -.
DR HSSP: P05091; LCW3.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldedh; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD.
FT NP_BIND 242 247 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 265 265 BY SIMILARITY.
FT ACT_SITE 299 299 BY SIMILARITY.
SQ SEQUENCE 497 AA; 53809 MW; 47CA353FFDBA50E8 CRC64;

Query Match 33.0%; Score 842; DB 1; Length 497;
Best Local Similarity 41.5%; Pred. No. 2.3e-58;
Matches 204; Conservative 84; Mismatches 183; Indels 20; Gaps 14;

QY 2 VLSPDEYKSE---LFINNEFYSSKGSERLTL--TNPWDESTVATDVHVAANAADVSAAV 55
DB 8 ITTPNGVKYEQPLGLFIDGEFV--KGAEGKTETINPSNEKPIVA-VHEATEKDQDVTAVA 64
QY 56 ASVQAVKCGPKWKFTCAQRAACMLKFAADLAENAEKLAERLESPTGRPVSMITHFDIPNM 115
DB 65 AAKAP-ESRWROVTFSTRGRMLTKLADLFEKDABILASIEALDNKKSITM-AHGDIAGA 122
QY 116 VSVRYAGWADKIAKTTFPDNGKPNM-RYEPNGVCGAGIASWATFLYGVNKAIPALAA 174
DB 123 AGCLRYVGWADKIHQOTIDTNSSETLNTYRHEPIGVCGQIIPNFWPLMIWKIGPAIAT 182
QY 175 GCSFIFKASEKSPGLVGLAPLAFARAGPPGVQFLTG-ARVTGEALASHMDIAKISFTR 233
DB 183 GNTWIKTAEQTLPSGLYAAANVKEAGIPAGVNVYISGFRVAGSAISHHMDIKVAFGT 242
QY 234 SVGGGKRAVQAATLLEGEK-PTIVNEAPLERQSGESAKDFSKFGQIWWPP 291
DB 243 STLVGRTIQLAAKSNLKKVTLLEGEKSPNVIFNDADIDNAISWNGGIFNHWQCCAC 302
QY 292 SCILLVQGNLAEKFGVHSGFGCCQRLQCNPLEPKRTHGPFVDSQTDRLVGNIDVGK 351
DB 303 SRILVQEG-IYDKF--IARLKERALQNKVG-DPFAKDTFQGPQVSOLOFDRIMEYIOHGK 358
QY 352 DT-AQLLTGVGRKDGKFAIEPTIFVNPKPCKSWPEEIPFPVLSIKTKTEFEAEIAND 410
DB 359 DAGATVAVGERHGTGTYFIQTFTVTDVTSMDKINQEEIFGVVTVOKFKDVEDAIKIGN 418
QY 411 DTTYGLASVIYTKSLNRLRVSSALETCGVSNFPPFIPETOTPFCKMGKSGSGRELGE 470
DB 419 STSGLAAGIHTKDVTTAIRVSNALRAGTVWVNSNLIQYQVFPFGFKESGIGRELGSYA 478
QY 471 LKAYLEPKTIN 481
DB 479 LENYQTIKAVH 489

RESULT 11
DHAS_CHICK STANDARD; PRT; 499 AA.
AC O93344.
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aldehyde dehydrogenase 1A2 (EC 1.2.1.3) (Retinaldehyde-specific
  dehydrogenase type 2) (RALDH(II)) (RALDH-2).
GN ALDH1A2 OR RALDH2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98394476; PubMed=9727493;
RA Sockanathan S., Jessell T.M.;
```


RL Alcohol 2:103-106(1985).
RN [9]
RC SEQUENCE OF 214-500 FROM N.A.
RC TISSUE=Liver;
RA MEDLINE=87279033; PubMed=3610592;
RA Agarwal D.P., Goedde H.W.;
RT "Human aldehyde dehydrogenase isozymes and alcohol sensitivity.";
RL Isozymes Curr. Top. Biol. Med. Res. 16:21-48(1987).
RN [10]
RN VARIANT LYS-504
RX MEDLINE=84119449; PubMed=6582480;
RA Yoshida A., Huang I.-Y., Ikawa M.;
RT "Molecular abnormality of an inactive aldehyde dehydrogenase variant
commonly found in Orientals.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:258-261(1984).
RN [11]
RN VARIANT LYS-496.
RX MEDLINE=96119362; PubMed=8561277;
RA Novorodovsky A., Tsai S.J., Goldfarb L., Peterson R., Long J.C.,
Goldman D.;
RT "Mitochondrial aldehyde dehydrogenase polymorphism in Asian and
American Indian populations: detection of new ALDH2 alleles.";
RL Alcohol. Clin. Exp. Res. 19:1105-1110(1995).
RN [12]
RX X-RAY CRYSTALLOGRAPHY (2.58 ANGSTROMS).
RA MEDLINE=20095857; PubMed=10631996;
RA Ni L., Zhou J., Hurley T.D., Weiner H.;
RT "Human liver mitochondrial aldehyde dehydrogenase: three-dimensional
structure and the restoration of solubility and activity of chimeric
forms.";
RL Protein Sci. 8:2784-2790(1999).
CC -|- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -|- PATHWAY: SECOND STEP IN ETHANOL UTILIZATION.
CC -|- SUBUNIT: HOMOTETRAMER.
CC -|- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -|- DISEASE: A VERY HIGH INCIDENCE OF ACUTE ALCOHOL INTOXICATION IN
ORIENTALS AND SOUTH AMERICAN INDIANS, AS COMPARED TO CAUCASIANS,
CAN BE DIRECTLY ATTRIBUTED TO AN ENZYMICALLY IMPAIRED ALDH
ISOZYME. THE INACTIVE VARIANT (ALLELE 2 OR ALDH2*2) IS DUE TO
SINGLE AMINO ACID EXCHANGE.
CC -|- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC -|- CAUTION: THE CONFLICTS BETWEEN THE SEQUENCE DETERMINED IN REF.1
AND REF.2 AND THAT IN REF.3 AND 4 ARE PROBABLY ALL DUE TO
FRAMESHIFT OR SEQUENCING ERRORS AS DESCRIBED IN REF.5 AND REF.6;
THE SEQUENCE DESCRIBED IN REF.9 DIFFERS FROM THAT SHOWN DUE TO TWO
FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M20456; AAA51693.1; -
DR EMBL; M20444; AAA51693.1; JOINED.
DR EMBL; M20445; AAA51693.1; JOINED.
DR EMBL; M20446; AAA51693.1; JOINED.
DR EMBL; M20447; AAA51693.1; JOINED.
DR EMBL; M20448; AAA51693.1; JOINED.
DR EMBL; M20449; AAA51693.1; JOINED.
DR EMBL; M20450; AAA51693.1; JOINED.
DR EMBL; M20451; AAA51693.1; JOINED.
DR EMBL; M20452; AAA51693.1; JOINED.
DR EMBL; M20453; AAA51693.1; JOINED.
DR EMBL; M20454; AAA51693.1; JOINED.
DR EMBL; M20455; AAA51693.1; JOINED.
DR EMBL; Y00109; CAA68290.1; -
DR EMBL; K03001; AAB59500.1; -
DR EMBL; M26760; AAA51694.1; -
DR EMBL; M54931; AAA62825.1; ALT_FRAME.
DR PIR; S00804; DEHUE2.

DR PIR; A29975; A29975.
DR PDB; 1CW3; 10-JAN-00.
DR MIN; 100650; -
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW OXidoreductase; NAD; Mitochondrion; Transit peptide; Polymorphism;
KW 3D-structure.
FT TRANSIT 1 17 MITOCHONDRION.
FT CHAIN 18 517 ALDEHYDE DEHYDROGENASE.
FT NP_BIND 262 267 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 285 285
FT ACT_SITE 319 319
FT VARIANT 337 337
FT VARIANT 496 496 E -> V (IN DBSNP:1062136).
FT VARIANT 504 504 /FTID=VAR_011869.
FT VARIANT 504 504 E -> K (IN ALDH2*3).
FT VARIANT 504 504 /FTID=VAR_011302.
FT CONFLICT 7 12 OF ENZYME ACTIVITY).
FT CONFLICT 7 26 REGPR -> ARAPP (IN REF. 5).
FT CONFLICT 18 85 AMPAWAPPLVSRHHPGRA (IN REF. 3).
FT CONFLICT 18 85 VKAARA -> REGRPG (IN REF. 3 AND 4).
FT CONFLICT 216 216 L -> S (IN REF. 9).
FT CONFLICT 218 218 A -> R (IN REF. 9).
FT CONFLICT 247 247 A -> P (IN REF. 9).
FT CONFLICT 380 380 E -> O (IN REF. 1).
FT CONFLICT 462 463 AL -> SP (IN REF. 9).
SQ SEQUENCE 517 AA; 56381 MW; E8F74044D285A00E CRC64;

Query Match 33.0%; Score 840.5; DB 1; Length 517;
Best Local Similarity 41.0%; Pred. No. 3.2e-58;
Matches 200; Conservative 86; Mismatches 187; Indels 15; Gaps 12;

QY 5 PDEYKSELFINNEFVSKGSERLTLNPWDESTVATDVHVANADVDVAASVQAVKKG 64
DB 32 PEVCNCFINNEHMDVSRKTFPTVNP-STGEVICOVAEGDKEDVKAVKAARAFGL 90
QY 65 -PWKKFTGAORAAACMLKAFADLAENAKELARLESPTGRPVSNITHFDIPNMVSFYYA 123
DB 91 SPWRMDASHRGLLNRLADLIERDRTYLAALFTLDNGKPYVISYLVLDLMVYKCLRYA 150
QY 124 GWADKIAGKTFPEDNGKPNM-RYEPMGVCGAGIASNNATFLYGVKIAAPALAGCSFIFKA 182
DB 151 GWADKYHGKTIPIGDFSTYRHEPVCGQIIPWNPFLMGAOKLGPALATGNVVMKV 210
QY 183 SEKSPGLVGLAPLAFAGPPGVQVLTGARVT-GEALASHMDIAKISFTRSVGGGRV 241
DB 211 AEQTPLTALYVANIKEAGPPGVNIVFGFGTAAAIASHEDVDKVAFTGSTEIGRVI 270
QY 242 KOATLKSNKRVLTLEGEK-PTIVFNEAPLERQSGESAKDFSKF---GOIWPVPSCLVQ 297
DB 271 QVAAGSSNLKRVLTLEGGKSPNIIIMSDADM-WAVEQA-HFALFFNOGCCGCCAGSRFVQ 328
QY 298 WGNLAERFHVGRHSGFGGCGORWLGONPLPKRTHGPFVDKSOYDRVLGNIDVGK-DTALQ 356
DB 329 -EDIYDEF--VERSVAKSRVVG-NPDSKTEQGPVDVETOPFKILGYINTKQEGAKL 384
QY 357 LTGVGRKDGKFAIEPTIFVNPFGSKIWFEIFGVPVLSIKTKFTEEAIEIANDTYGL 416
DB 385 LCGGIIAADRGYFIQPTVFGVDGDMGTIAKEEIFGVMOILKFKTFIEEVGRANNSTVGL 444
QY 417 ASVIYTKSLNRGRVSSALETGGVSNFPFIPETQTTPFGMKQSGRELGEGLKAYLE 476
DB 445 AAATFKDLKANYLSQALQAGTVWVNCYDVFGAOSPFGGYKMSGRELGEGLQAVTE 504
QY 477 PRTNIHV 484
DB 505 VKTVYKV 512

```

RESULT 13
DHAE_MACPR STANDARD; PRT; 501 AA.
AC Q29490;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Aldehyde dehydrogenase, cytosolic 1 (EC 1.2.1.3) (ALDH class 1) (ETA-
DE Crystallin).
OS Macroscelides proboscideus (Short-eared elephant shrew).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Macroscelidea; Macroscelididae; Macroscelides.
OX NCBI_TaxID=29082;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RX MEDLINE=96279083; PubMed=8663049;
RA Graham C., Hodin J., Wistow G.;
RT "A retinaldehyde dehydrogenase as a structural protein in a mammalian
RL eye lens. Gene recruitment of eta-crystallin.",
CC J. Biol. Chem. 271:15623-15628(1996).
CC -|- FUNCTION: MAJOR COMPONENT OF THE EYE OF ELEPHANT SHREWS, WHICH IN
CC CONTRAST TO OTHER MAMMALS, POSSESSES BOTH A LENS- AND A NON-LENS
CC CLASS-1 ALDEHYDE DEHYDROGENASE 1. THIS EYE-SPECIFIC FORM IS A
CC STRUCTURAL PROTEIN OF THE LENS AND, IN OTHER PART OF THE EYE,
CC SERVES AS THE MAJOR FORM OF ALDH1. CAN CONVERT/OXIDIZES
CC RETINALDEHYDE TO RETINOIC ACID.
CC -|- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -|- PATHWAY: SECOND STEP IN ETHANOL UTILIZATION (BY SIMILARITY).
CC -|- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (BY SIMILARITY).
CC -|- TISSUE SPECIFICITY: EYE-SPECIFIC, WITH VERY HIGH EXPRESSION IN THE
CC LENS.
CC -|- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U03906; AAC48588.1; -
CC HSSP; P51977; 1BX5.
CC InterPro; IPR002086; Aldehyde_dehydr.
CC Pfam; PF00171; aldehyd; 1.
CC PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
CC PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; FALSE_NEG.
CC
CC Oxidoreductase; NAD.
CC NP_BIND 246 251 NAD (ADP PART) (BY SIMILARITY).
CC ACT_SITE 269 269 POTENTIAL.
CC ACT_SITE 303 303 POTENTIAL.
CC SEQUENCE 501 AA; 54624 MW; 48C15C682665D0C0 CRC64;

```

```

Query Match 32.9%; Score 838.5; DB 1; Length 501;
Best Local Similarity 40.6%; Pred. No. 4.3e-58;
Matches 196; Conservative 92; Mismatches 182; Indels 13; Gaps 10;

```

```

QY 188 LGVLGLAPLFAAGPPPGVQFLTGARVT-CEALASHMDTAKISFTSRVCGGGRVAKQNTL 246
DB 200 LTALHVASLIKKEAGPPPGVNVIVPGYPTAGAAISSHMDVDKVAFTGSTEVGKMIQAAA 259
QY 247 KSNMKRVLTLEGEK-PTIVFNEAPLEROSGESAKD--FSKFGQIWPFPSCLLVQWGNLAE 303
DB 260 KSNLKRVTLELGAKNPCIVFADALD-SAVEFAHQGVFTNQGGSCIAASKLFVE-ETIYD 317
QY 304 KFGVPHSGFCGCRWLGNPLKRGTHGPFVDKSOYDRVLGNIDVG-KDTAOLLTGVR 362
DB 318 EF---VORSEVERAKKIVFGNPLTPGVNHGPNQINAKQINKIMELIESKKGAKLEGGGP 374
QY 363 KGDGFAIEPTIFVNPXKPGSKIMFEEIPGVLISIKTKTEEEAIEIANDTYGLASVIYT 422
DB 375 WGNKGYFTQPTIFSNVTDMMRIAKETIEGVPQQIMKPKSLDEVIKRANNTYYGLVAGVET 434
QY 423 KSLNGLRLVSSSALETGYSINFPFIPETQTPFGGKMGSGGREGELGKAYLEPKTINI 482
DB 435 KOLDKRAVTVSSALQAGTVWVNCYLAASAQSPAGGFKMGGHGREMGEYGIHEYTEVKTVM 494
QY 483 HVN 485
DB 495 KIS 497

```

RESULT 14

```

DHAE_CHICK STANDARD; PRT; 509 AA.
AC P27463;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aldehyde dehydrogenase 1A1 (EC 1.2.1.3) (Aldehyde dehydrogenase,
DE cytosolic) (ALDH class 1).
GN ALDH1A1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryonic retina;
RX MEDLINE=92217647; PubMed=1559558;
RA Godbout R.;
RT "High levels of aldehyde dehydrogenase transcripts in the
RT undifferentiated chick retina.",
RL Exp. Eye Res. 54:297-305(1992).
CC -|- FUNCTION: BINDS FREE RETINAL AND CELLULAR RETINOL-BINDING PROTEIN-
CC BOUND RETINAL. CAN CONVERT/OXIDIZES RETINALDEHYDE TO RETINOIC ACID
CC (BY SIMILARITY).
CC -|- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -|- PATHWAY: SECOND STEP IN ETHANOL UTILIZATION.
CC -|- SUBUNIT: HOMOTETRAMER.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
CC -|- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X58869; CAA41679.1; -
CC PIR; S14629; S14629.
CC HSSP; P51977; 1BX5.
CC InterPro; IPR002086; Aldehyde_dehydr.
CC Pfam; PF00171; aldehyd; 1.
CC PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
CC PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.

```

```
KW Oxidoreductase; NAD.
FT NP_BIND 254 259 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 277 277 BY SIMILARITY.
FT ACT_SITE 311 311 BY SIMILARITY.
SQ SEQUENCE 509 AA; 55809 MW; 7771181FA2F05DA9 CRC64;

Query Match 32.98; Score 838.5; DB 1; Length 509;
Best Local Similarity 39.88; Pred. No. 4.4e-58;
Matches 192; Conservative 87; Mismatches 190; Indels 13; Gaps 8;

QY 10 SELFNNFVSSKSGSERLTLPNWDSTVATDVHANAADVSAVAASQVAKKG-PWKK 68
DB 29 TKIFINNEHDSVSGKFEVENPANEKIC-EVAEGKADIDKAVKAARAFELGSPWRT 87
QY 69 FTGAORAAACMLKADLAENAEKLAERLESPTGRPSMITHFDIPNMVSVFRYYAGWADK 128
DB 88 MDASERGRLLKLADLVERDRLTATMEADGGKLFSTAYLMDLGACIKTIRTCAGWADK 147
QY 129 IAGKTFPED-NGKPNRYEPMGVCAGIASMNATFLYGVGKIAPALAAAGCSFIFKASEKSP 187
DB 148 IHGRTPVMDGNFTFTRHEPVGVCGQIIPNPNFLPMFIWKIAPALCCGNTVVVKPAEQTP 207
QY 188 LGVLGLAPLFAFAGPPGVQFLGARGVT-GEALASHMDIAKISFTRSVGGGRAVQOATL 246
DB 208 LSALYMGSLIKEAGFPVGVNIVPGFGPTAGAAISHHMDIKVSFTGSTEVGKLKEAAG 267
QY 247 KSNMKRYTLELGEK-PTIVFNEAPLERQSGESAKDFSGFQGIWVPPSCLLVQVGNLAEK- 304
DB 268 KTNLKVITLELGGKSPNIFADADL-----DEAAEFHIGLFYHOGGCCAGTSRIFVEEP 322
QY 305 -FHGVRHSGFGCCORWLGNLPLEKRTGHPVDKSOYDVLGNIDVG-KDTAOLLTCGVGR 362
DB 323 IYDFEVRSIERAKRYTLGDPQLPCVOGQIDKEQFQKILDLIESGKKGAKLECGGSP 382
QY 363 KDKGFAIEPTIFVNPXPGSKINFEIIFGVPVLSIKTEKTEEEAIEANDTTTGLASVIIT 422
DB 383 WGNKYIFQPVAFVSVDMDRIAKEEIFGPVQIQMKTEIDIVIKRANNTTYGLAAAVFT 442
QY 423 KSLNGLRVSSALTEGVGVSINFPFIPETPTPFQGMKSGSGRELGBEGLKAYLEPKTINI 482
DB 443 KDIDKALFASALQAGTWWNCYSAFSAQCPFGFGKMSGNGRELGEVGLQYEYEVKTVTI 502
QY 483 HV 484
DB 503 KI 504

RESULT 15
DH2A_MOUSE STANDARD; PRT; 499 AA.
AC Q62148;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aldehyde dehydrogenase 1A2 (EC 1.2.1.3) (Retinaldehyde-specific
DE dehydrogenase type 2) (RALDH1(1)) (RALDH-2).
GN ALDH1A2 OR ALDH1A7 OR RALDH2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/HE;
RA MEDLINE=96390857; PubMed=8797830;
RA Zhao D., McCafferty P., Ivins K.J., Neve R.L., Hogan P., Chin W.W.,
RA Draeger U.C.;
RT "Molecular identification of a major retinoic-acid-synthesizing
RT enzyme, a retinaldehyde-specific dehydrogenase."
RL Eur. J. Biochem. 240:15-22(1996).
CC -!- FUNCTION: RECOGNIZES AS SUBSTRATES FREE RETINAL AND CELLULAR
CC RETINOL-BINDING PROTEIN-BOUND RETINAL. DOES METABOLIZE OCTANAL AND
```

```
CC DECANAL BUT DOES NOT METABOLIZE CITRAL, BENZALDEHYDE, ACETALDEHYDE
CC AND PROPANAL EFFICIENTLY (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -!- PATHWAY: RETINOIC ACID BIOGENESIS.
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X99273; CAA67666.1; -.
CC HSP; P51977; IBXS.
CC MGD; MGI:107928; Aldh1a2.
CC InterPro; IPR002086; Aldehyde_dehydr.
CC Pfam; PF00171; aldedh; 1.
CC PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
CC PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
CC Oxidoreductase; NAD.
CC NP_BIND 244 249 NAD (ADP PART) (BY SIMILARITY).
CC ACT_SITE 267 267 POTENTIAL.
CC ACT_SITE 301 301 POTENTIAL.
CC SEQUENCE 499 AA; 54725 MW; E0000596A30B0B7A CRC64;
CC -----
```

Query Match 32.88; Score 836.5; DB 1; Length 499;
Best Local Similarity 40.7%; Pred. No. 6.2e-58;
Matches 196; Conservative 83; Mismatches 191; Indels 11; Gaps 10;

```
QY 10 SELFNNFVSSKSGSERLTLPNWDSTVATDVHANAADVSAVAASQVAKKG-PWKK 68
DB 19 TKIFINNEHDSVSGKFEVENPANEKIC-EVAEGKADIDKAVKAARAFELGSPWRT 77
QY 69 FTGAORAAACMLKADLAENAEKLAERLESPTGRPSMITHFDIPNMVSVFRYYAGWADK 128
DB 78 MDASERGRLLKLADLVERDRLTATMEADGGKLFSTAYLMDLGACIKTIRTCAGWADK 137
QY 129 IAGKTFPEDNGKPNW-RYPEMVGVCAGIASMNATFLYGVGKIAPALAAAGCSFIFKASEKSP 187
DB 138 IHGNTIPVDGDYFTFTRHEPVGVCGQIIPNPNFLPMFIWKIAPALCCGNTVVVKPAEQTP 197
QY 188 LGVLGLAPLFAFAGPPGVQFLGARGVT-GEALASHMDIAKISFTRSVGGGRAVQOATL 246
DB 198 LSALYMGSLIKEAGFPVGVNIVPGFGPTAGAAISHHMDIKVSFTGSTEVGKLKEAAG 257
QY 247 KSNMKRYTLELGEK-PTIVFNEAPLERQSGESAKD-FSKFGQIWPVPPSCLLVQVGNLAEK 304
DB 258 RSNLKVITLELGGKSPNIFADADLQYAEQAHOGVFFNQGCCAGTSRIFVE-ESIYEE 316
QY 305 FHGVRHSGFGCCORWLGNLPLEKRTGHPVDKSOYDVLGNIDVG-KDTAOLLTCGVGR 363
DB 317 F--VKRSVERAKRRIVG-SPFDPTEGPOIDKQKYNKVLLELIQSGVAGAKLECGGKL 373
QY 364 GDKGFAIEPTIFVNPXPGSKINFEIIFGVPVLSIKTEKTEEEAIEANDTTTGLASVIITK 423
DB 374 GRKGFFTEPTVFSNVTDMDRIAKEEIFGPVQIEKREKTMDEVIERANNSDFGLVAAVFTN 433
QY 424 SLNRGLRVSSALTEGVGVSINFPFIPETPTPFQGMKSGSGRELGBEGLKAYLEPKTINI 483
DB 434 DINKALMVSSAMQGVWVWNCYNALNAQSPFGCFKMSGNGREMGEGFLKEYSVATVTVK 493
QY 484 V 484
DB 494 I 494
```

Search completed: April 27, 2002, 07:59:39

Job time: 745 sec

| | | | |
|----|-----|---|-----|
| QY | 424 | SLNRGLRVSALETGGVSVINFPFIPETQTPFGMGKQSGRGELGKAYLEPKTINH | 483 |
| Db | 435 | DLQKAITVSAAALQAGTVWVNCYGVVSAQCPSFGFKMCSNGRMGVGFHEYTEVKTVIVK | 494 |
| QY | 484 | VN | 485 |
| Db | 495 | IS | 496 |

N:Alternate names: aldehyde dehydrogenase E1
C:Species: Homo sapiens (man)
C:Date: 28-May-1986 #sequence_revision 31-Dec-1993 #text_change 08-Dec-2000
C:Accession: A33371; I51846; I39431; JC5554; A00363; A61010; A26817
R:Hsu, L.C.; Chang, W.C.; Yoshida, A.
Genomics 5, 857-865, 1989
A>Title: Genomic structure of the human cytosolic aldehyde dehydrogenase gene.
A:Reference number: A33371; MUID:90077427
A:Accession: A33371
A:Molecule type: DNA
A:Residues: 1-501 <H5U>
A:Cross-references: GB:M31994; NID:q178370; PIDN:AAA51692.1; PID:q178372
R:Yoshida, A.; Hsu, L.C.; Yanagawa, Y.
Adv. Exp. Med. Biol. 328, 37-44, 1993
A>Title: Biological role of human cytosolic aldehyde dehydrogenase 1: hormonal response.
A:Reference number: I51846; MUID:93263033
A:Accession: I51846
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-6 <YOS1>
A:Cross-references: GB:S61235; NID:q300399; PIDN:AAD13925.1; PID:q4261625
R:Yoshida, A.; Ikawa, M.; Hsu, L.C.; Tani, K.
Alcohol 2, 103-106, 1985
A>Title: Molecular abnormality and cDNA cloning of human aldehyde dehydrogenases.
A:Reference number: I39431; MUID:85252089
A:Accession: I39431
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-163-501 <YOS2>
A:Cross-references: GB:M26761; NID:q178393; PIDN:AAA35518.1; PID:q178394
R:Kathmann, E.C.; Lipsky, J.J.
Biochem. Biophys. Res. Commun. 236, 527-531, 1997
A>Title: Cloning of a cDNA encoding a constitutively expressed rat liver cytosolic aldehyde dehydrogenase.
A:Reference number: JC5553; MUID:97382470
A:Accession: JC5554
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-120,'S',122-501 <KAT>
A:Cross-references: DDBJ:AF003341; NID:q2183298; PIDN:AAC51652.1; PID:q2183299
A:Experimental source: liver
R:Hempel, J.; von Bahr-Lindstrom, H.; Jorvall, H.
Eur. J. Biochem. 141, 21-35, 1984
A>Title: Aldehyde dehydrogenase from human liver. Primary structure of the cytoplasmic isozyme.
A:Reference number: A00363; MUID:84208019
A:Accession: A00363
A:Molecule type: protein
A:Residues: 2-501 <HEM>
R:Agarwal, D.P.; Cohn, P.; Goedde, H.W.; Hempel, J.
Enzyme 42, 47-52, 1989
A>Title: Aldehyde dehydrogenase from human erythrocytes: structural relationship to the enzyme from human liver.
A:Reference number: A61010; MUID:89377753
A:Accession: A61010
A:Molecule type: protein
A:Residues: 23-27;79-85;101-107;114-128;132-142;144-154;309-319;421-434;477-483 <AGA>
R:Experimental source: erythrocytes
R:Abriola, D.P.; Figels, R.; Stein, S.; MacKerell Jr., A.D.; Pietruszko, R.
Biochemistry 26, 5679-5684, 1987
A>Title: Active site of human liver aldehyde dehydrogenase.
A:Reference number: A26817; MUID:88050861
A:Accession: A26817
A:Molecule type: protein
A:Residues: 266-273 <ABR>
A>Note: the active site Glu was identified by suicide inhibition with bromoacetophenone
C:Genetics:
A:Gene: GDB:ALDH1
A:Cross-references: GDB:119667; OMIM:100640
A:Map position: 9q21-9q21
C:Complex: homotetramer
C:Function:
A:Description: catalyzes oxidation of an aldehyde to an acid using NAD+ and water
A>Note: enzymes with this activity are involved in diverse metabolic pathways in various tissues
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homolog
C:Keywords: acetylated amino end; alcohol metabolism; cytosol; homotetramer; liver; NAD+

F;2-501/product: aldehyde dehydrogenase (NAD+) 1 #status experimental <MAT>
F;59-323/domain: aldehyde dehydrogenase homology <ALDD>
F;2/Modified site: acetylated amino end (Ser) (in mature form) #status experimental
F;269,303/Active site: Glu, Cys #status experimental
F;456/Binding site: NAD (Cys) #status experimental

Query Match 33.1% Score 843.5; DB 1; Length 501;
Best Local Similarity 39.8%; Pred. No. 2.2e-58;
Matches 192; Conservative 90; Mismatches 188; Indels 13; Gaps 8;

QY 10 SELFNNFVSSKSGSERLTLTPWDESTVNTDVHANAADVDVAASVQAVKKG-PWKX 68
Db 21 TKIFINNEHDSVSGKFFVFNPAETEEELC-OVEEGREDVDKAVKAARQAFQIGSPWT 79
QY 69 FTGAQAAACMLKPADLAENAKNAELARLESPLTPGPRVSMITHFDIPNMVSVERYAGNADK 128
Db 80 MDASERGRLLYKLADLIERDRLLATMESMNGGKLYSNAYLNDLAGCIKTURLYACAGWADK 139
QY 129 IAGKTFPED-NGKPNWRYPGMCVAGIASWNATLYVGWKIAPALAAAGCSPIFKASEKSP 187
Db 140 IQGRTTIDGNFTYTRHEPIGVCGQIIPWNPFLVWLTKICPALSCTNTVVVPAEQTP 199
QY 188 LGVLGLAPLFAEAGPPGVPVQVLTGARVT-GEALASHMDIAKISFTRSVGGGRAVKQATL 246
Db 200 L7ALHVASLKEAGFPVGVNIPVGYGTAGAAATSSHMDIKVAFPTGSTEYKLIKEAAG 259
QY 247 KSNMKRVTLLEGEK-PTIVNEAPLERQSGESAKDFSGFGLIWPVPPSCLLVOMGNLAEK- 304
Db 260 KSNLKRVTLELGGKSPCIVLADADL-----DNAVEFAHGVFYHOGQCCIAASRIFVEES 314
QY 305 -FHGVYHSGFGGCGQWLGONPLPKRTHGPFVVDKSOYDRVLGNIDVG-KDTAQLLTGVGR 362
Db 315 IYDEFYRSVERAKKYLGNLTPGVTQPGQIDKQYDKILDLIESGKKEGAKLECCGGP 374
QY 363 KDGKGAIPETIFVNPKPSKIWFEEIFGPVLSIKTFTEEAETIANDTTYGLASVIYT 422
Db 375 WGNKGVYQPTVFSNVTDEMIAKEEIEFGVQOIMKFKSLDDVIKRANNTYGLSAGVFT 434
QY 423 KSLNRLRVSSALETGGVGINPFPETQTPFGGMKQSGSRELGEGELKAYLEPKYINI 482
Db 435 KDIDKAITISSALQATVWVNCYGVVSAQCPFGFGKMSGNGRELGEYGFHEYTEYKTVTV 494
QY 483 HVN 485
Db 495 KIS 497

RESULT 9
JC4524
aldehyde dehydrogenase (NAD(P)+) (EC 1.2.1.5) - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 08-Feb-1996 #sequence_revision 19-Apr-1996 #text_change 24-Nov-1999
C:Accession: JC4524; PC4119
R:Bhat, P.V.; Labrecque, J.; Boutin, J.M.; Lacroix, A.; Yoshida, A.
Gene 166, 303-306, 1995
A>Title: Cloning of a cDNA encoding rat aldehyde dehydrogenase with high activity for
A:Reference number: JC4524; MUID:96125208
A:Accession: JC4524
A:Molecule type: mRNA
A:Residues: 1-501 <BHA>
A:Cross-references: GB:L42009; NID:q1256958; PIDN:AAA96657.1; PID:q974168
A:Accession: PC4119
A:Molecule type: protein
A:Residues: 2-19;80-84;86;88-91;96-99;101-108;110-121;205-209;213-215;218-219;222;22
A:Experimental source: Kidney
C:Genetics:
A:Gene: aldH
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: acetylated amino end; kidney; NAD; oxidoreductase
F;2-501/product: aldehyde dehydrogenase #status predicted <MAT>
F;59-323/domain: aldehyde dehydrogenase homology <ALDD>
F;2/Modified site: acetylated amino end (Ser) (in mature form) #status predicted

Db 32 PEVFCNQIFLNNEWHDAVSRKFTPTVNP-STGEVTCQVAEGDKEDVDKAVKAAAFOLG 90
QY 65 -PKKFTGAQACMLKFLADLAEKNAELARLESIPTRGVPVSMITHFDIPNNVSVFYYA 123
Db 91 SPWRMDASHRGLINLADLIERDRTYLALETLDNGKPPYVSYLVLDLDMVKLCRYA 150
QY 124 GWADKTAGKTFPDNKGPNW-RYEPMGVCAGIASNATFLYVGVKIAAPALAAAGCSFIPKA 182
Db 151 GWADKHGKTIPIIDGDFSTREHPVGVCGQIIPWNPFLMQAWKLGFPALATGNVVMKV 210
QY 183 SEKSPGLVGLAPLFAEAGFPVGVQVLTGARVT-GEALASHMDIAKISFTRSVGGGNAV 241
Db 211 AEQTPLTALVYANLKEAGFPVGVNIVPFGPTAGAAIASHEDVDKVAFTGSTEIGRVI 270
QY 242 KQATLKSMMKRVLELGEK-PTIVFNEAPLERQSGESAKDFSKF---GQIWWPPSCLLVQ 297
Db 271 QVAAAGSNLKVTLLEGGKSPNIMSADMD-WAVEQA-HFALFFNQGCCAGSRTEVQ 328
QY 298 WGNLAERHGVHSGFGGCGRMGONPLEPKRTHGPFVDKSOYDRVLGNIDVGK-DTAQL 356
Db 329 -EDIYDEF--VERSARAKSRVVG-NPDSKTEQGPQVDETQFKILGYINTKGEGAKL 384
QY 357 LTGVGRKDGKFAIEPTIFVNPKPGSKIWFEEIFGVPVLSIKTFKTEEAIEIANDTTVGL 416
Db 385 LCGGGAADRGYPIQPTVFEGVDGDMTIAKEEIFGVPVQILKFKTIEEVVGRANNSTYGL 444
QY 417 ASVYTKSLNKLRLVSSALETGGVSNFPFIPETQTPFGGKMGOSGRELGEGLKAYLE 476
Db 445 AAAVFTKDLKANYLSQALQAGTVWVNCYDVGASFPFGYKMGSGRELGEYGLQAYTE 504
QY 477 PNTNIHV 484
Db 505 VKTIVTKV 512

RESULT 11

S14629
aldehyde dehydrogenase (NAD+) (EC 1.2.1.3) 1, cytosolic - chicken
N:Alternate names: Gallus gallus (chicken)
C:Species: Gallus gallus (chicken)
C:Date: 21-Nov-1993 #sequence_revision 31-Jan-1997 #text_change 11-Jun-1999
C:Accession: S14629
R:Godbout, R.
submitted to the EMBL Data Library, April 1991
A:Description: High levels of aldehyde dehydrogenase transcripts in the undifferentiated
A:Reference number: S14629
A:Accession: S14629
A:Molecule type: mRNA
A:Residues: 1-509 <GOD>
A:Cross-references: EMBL:X58869; NID:g63032; PIDN:CAA41679.1; PID:g63033
A:Experimental source: retina
C:Complex: homotetramer
C:Function:
A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD+ and water
A:Pathway: ethanol catabolism
A:Note: enzymes with this activity are involved in diverse metabolic pathways in various
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: alcohol metabolism; cytosol; homotetramer; liver; NAD; oxidoreductase
F:67-331/Domain: aldehyde dehydrogenase homology <ALDD>
F:277,311/Active site: Glu, Cys #status predicted
F:464/Binding site: NAD (Cys) #status predicted

Query Match 32.9%; Score 838.5; DB 1; Length 509;
Best Local Similarity 39.8%; Pred. No. 5.5e-58;
Matches 192; Conservative 87; Mismatches 190; Indels 13; Gaps 8;

QY 10 SELFIPNNEFVSSKGSERLTLTNPWDESTVATDVHANAADVDSDVAASVQAVKKG-PWK 68
Db 29 TKIFINNEHDSVSGKKFEYFNANEKIC-EVAEGDKADIDKAVKAKAFELGSPWRT 87
QY 69 FTGAQACMLKPADLAEKNAELARLESIPTRGVPVSMITHFDIPNNVSVFYYAGWADK 128

Db 88 MDASERGRLLNKLAIDLIVERDRRLTLATMEIDGGKLFSTAYLMDLGACIKTIRYCAGWADK 147
QY 129 IAGKTFPED-NGKPNRREYPMGVCAGIASNATFLYVGVKIAAPALAAAGCSFIFKASEKSP 187
Db 148 IHGRTVPMGDNFTTTRHEPVGCGQIIPWNPFLVMIWKIAPALCCGNTVVKPAEQTP 207
QY 188 LGVLGLAPLFAEAGFPVGVQVLTGARVT-GEALASHMDIAKISFTRSVGGGNAVQKALT 246
Db 208 LSALYMGSLIKEAGFPVGVNIVPFGPTAGAAISHMDIDKVSFTGSTEVGKLLKEAG 267
QY 247 KSNKRVTLLELGEK-PTIVFNEAPLERQSGESAKDFSKFGQIWWPPSCLLVQWGLAEK- 304
Db 268 KTNLKRVTLELGGKSPNIPADAL-----DEAAEFAGHIGLEYHQGCCAGSRIFVEEP 322
QY 305 -FHGVRHSGFGGCGRMGONPLEPKRTHGPFVDKSOYDRVLGNIDVG-KDTAQLLTGVGR 362
Db 323 IYDEFVRSITERAKRTLTGDPLLPGVQGPQIDKGFQKILDLIESGKKEGAKLCEGGGP 382
QY 363 KGDGFAIEPTIFVNPKPGSKIWFEEIFGVPVLSIKTFKTEEAIEIANDTTVGLASVIYT 422
Db 383 WGNKGVFIQPTVFESNVTDDMRIAKEEIFGVPVQIIMFKTIDSVIKRANNTVGLAAAVET 442
QY 423 KSLNKLRLVSSALETGGVSNFPFIPETQTPFGGKMGOSGRELGEGLKAYLEKPTNI 482
Db 443 KDDKALTAFASALQAGTVWVNCYSAFSAOCFFGFKMNGRELGEYGLQETEYKVTI 502
QY 483 HV 484
Db 503 KI 504

RESULT 12

S74224
aldehyde dehydrogenase (NAD+) (EC 1.2.1.3) 2 - mouse
N:Alternate names: retinaldehyde-specific dehydrogenase
C:Species: Mus musculus (house mouse)
C:Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 20-Jun-2000
C:Accession: S74224
R:Zhao, D.; McCaffery, P.; Ivins, K.J.; Neve, R.L.; Hogan, P.; Chin, W.W.; Draeger, U.
Eur. J. Biochem. 240, 15-22, 1996
A:Title: Molecular identification of a major retinoic-acid-synthesizing enzyme, a ret
A:Reference number: S74224; MUID:96390857
A:Accession: S74224
A:Molecule type: mRNA
A:Residues: 1-499 <ZHA>
A:Cross-references: EMBL:X99273; NID:g1430868; PIDN:CAA67666.1; PID:g1430869
A:Experimental source: strain C3H/He; cell type embryonal carcinoma; cell line P19 te
C:Genetics:
A:Gene: RALDH-2
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: NAD; oxidoreductase
F:57-321/Domain: aldehyde dehydrogenase homology <ALDD>
F:193-273/Domain: NAD binding #status predicted <NAD>
F:267,301/Active site: Glu, Cys #status predicted
F:454/Binding site: NAD (Cys) #status predicted

Query Match 32.8%; Score 836.5; DB 2; Length 499;
Best Local Similarity 40.7%; Pred. No. 7.7e-58;
Matches 196; Conservative 83; Mismatches 191; Indels 11; Gaps 10;

QY 10 SELFIPNNEFVSSKGSERLTLTNPWDESTVATDVHANAADVDSDVAASVQAVKKG-PWK 68
Db 19 TKIFINNEHDSVSGRVPVPCNPATGEQVC-EVOEADKVIDKAVQARLAFSLGSVMRR 77
QY 69 FTGAQACMLKPADLAEKNAELARLESIPTRGVPVSMITHFDIPNNVSVFYYAGWADK 128
Db 78 MDASERGRLLNKLAIDLIVERDRATLWESLNGGKPKFLOAFYIDLOGVIKTRYYAGWADK 137
QY 129 IAGKTFPEDNKGPNW-RYEPMGVCAGIASNATFLYVGVKIAAPALAAAGCSFIFKASEKSP 187
Db 138 IHGRTIPVDGDTFTTRHEPVGCGQIIPWNPFLMFTWKIAPALCCGNTVVVVKPAEQTP 197

A:Molecule type: protein
A:Residues: 327-340 <DIW>
R:Jeng, J.; Weiner, B.
Arch. Biochem. Biophys. 289, 214-222, 1991
A:Title: Purification and characterization of catalytically active precursor of rat liver
A:Reference number: S17492; MUID:91378548
A:Accession: S17492
A:Molecule type: protein
A:Residues: 1-19 <JEN>
C:Genetics:
A:Genome: nuclear
C:Complex: homotetramer
C:Function:
A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD+ and water
A:Pathway: ethanol catabolism
A:Note: enzymes with this activity are involved in diverse metabolic pathways in various
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: alcohol metabolism; homotetramer; liver; mitochondrion; NAD; oxidoreductase
F:1-19/Domain: transit peptide (mitochondrion) #status predicted <NTP>
F:1-19/Product: aldehyde dehydrogenase (NAD+) 2 #status predicted <MAT>
F:77-341/Domain: aldehyde dehydrogenase homology <ALDD>
F:287,321/Active site: Glu, Cys #status predicted
F:474/Binding site: NAD (Cys) #status predicted

Query Match 32.7%; Score 833.5; DB 1: Length 519;
Best Local Similarity 40.8%; Pred. No. 1.4e-57;
Matches 199; Conservative 84; Mismatches 190; Indels 15; Gaps 12;
QY 5 PREYKSELPINNEFVSSKSGSERLTLPNWDSESTVATDVHVAADVDVAASVQAVKKG 64
DB 34 PEVFCNQIFINNEHDAVSKRTFTVNP-STGEVICQVAEGNKEDVDKAVKAAQAFQLG 92
QY 65 -PWKFTGAQRAACMLKADLAENAEKLALESPTGPRVSMITHFDIPNMVSVFRYYA 123
DB 93 SPWRMDASDRGLLYRLADLIERDRLTLALETLDNGRPYVISYLVLDLWVLCRLYYA 152
QY 124 GWADIAKTPEDNGKPNM-RYEPMGVCAGIASWNATFLYGVGKIAPALAAAGCSFTFKA 182
DB 153 GWADYHKGKTPIDGDFSYTRHEPVGVCGQIIPNFPILMLKAWKLGALATGNVVMKV 212
QY 183 SEKSLGVGLAPLFAEAGPPGVVQPLTGARVT-GEALASHMDIAKISFTRSVGGGRV 241
DB 213 ASQTLTALYVANLIKAEAGPPGVVNIIVPGFPTAGAAIASHEDVDKVAFTGSTEVEGHLI 272
QY 242 KQATLKSNNKRVTLLEGK-PTIVFNEAPLERQSGESAKDFSKF---QQIWWPPSCLLVQ 297
DB 273 QVAAGSSNLKRVTLLEGKSPNIIMSDADM-WAVEQA-HFALFPNQSCCCAGSRFTVQ 330
QY 298 WGNLAKEFWRHSGFGGCGORWLGONPLEPKRTHGPFVDSKSOYDRVLGNIDVG-KDTAQL 356
DB 331 -EDVDEF--VERSVARAKSRVVG-NPFDSTRTEQGFQVDETQFKILGTIKSGOQEGAKL 386
QY 357 LFGVGRKGDGFAIEPTIFVNPKPGSKIMPEEIFGPVLSIKTFTEEAIEANDTTYGL 416
DB 387 LCCGGAADRGYFIQPTVFGVDKDGWTIAKEEIFGPMOILFKTIEVVGGRANNKYGL 446
QY 417 ASVIVTKSLNGLRVSALTEGVSINFPETQTPFGGMKQSGSGRELGEGLKAYLE 476
DB 447 AAATPKDLDRANLYSQLAQGTWVINCVDVFGAQSPFGGKMSGSGRELGEYGLQAYTE 506
QY 477 PKTINHV 484
DB 507 VKTIVTKV 514

RESULT 15
S14752
aldehyde dehydrogenase (NAD+) (EC 1.2.1.3), cytosolic - sheep
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 19-Mar-1997 #sequence_revision 13-Sep-1998 #text_change 21-Jan-2000
C:Accession: S78582; S14752
R:Weedle, J.W.

submitted to the EMBL Data Library, July 1994
A:Reference number: S78582
A:Accession: S78582
A:Molecule type: mRNA
A:Residues: 1-501 <WWE>
A:Cross-references: EMBL:U12761; NID:9527681; PIDN:AAA85435.1; PID:9527682
A:Experimental source: liver
A:Kitson, T.M.; Hill, J.P.; Midwinter, G.G.
Biochem. J. 275, 207-210, 1991
A:Title: Identification of a catalytically essential nucleophilic residue in sheep l
A:Reference number: S14752; MUID:91207293
A:Accession: S14752
A:Molecule type: protein
A:Residues: 298-308 <KIT>
A:Experimental source: liver
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: NAD; oxidoreductase
F:59-323/Domain: aldehyde dehydrogenase homology <ALD>
F:302/Active site: Cys #status experimental

Query Match 32.6%; Score 830.5; DB 2: Length 501;
Best Local Similarity 40.8%; Pred. No. 2.3e-57;
Matches 197; Conservative 91; Mismatches 182; Indels 13; Gaps 11;

QY 10 SELFNNFVSSKSGSERLTLPNWDSESTVATDVHVAADVDVAASVQAVKKG-PWKK 68
DB 21 TKIPINNEHSSVSGKKFPFVFNPAEELK-EVERGDKEDVDKAAKAAQAFQIGSPWRT 79
QY 69 FTGAQRAACMLKADLAENAEKLALESPTGPRVSMITHFDIPNMVSVFRYYAGWAK 128
DB 80 MDASERGLNKLADLIERDRLTLALETLDNGRPYVISYLVLDLWVLCRLYYA 139
QY 129 IAGTTEPED-NGKPNRYEPMGVGVCAGIASWNATFLYGVGKIAPALAAAGCSFIFKASEKSP 187
DB 140 IQGRTIPMDGNFTYTRSEPVGVCGQIIPNFPILMLKAWKLGALATGNVVMKV 199
QY 188 LGVLGLAPLFAEAGPPGVVQPLTGARVT-GEALASHMDIAKISFTRSVGGGRVQKQATL 246
DB 200 LTALHMGSLIKEAGFPVGVVNIIVPGYGTAGAAISSHMDVDKVAFTGSTEVEGLINEAAG 259
QY 247 KSNMKRVTLLEGK-PTIVFNEAPLERQSGESAKD--FSKFGQIWWPPSCLLVQWGNLAE 303
DB 260 KSNLKRVSLELGKSPCIIVFADADLD-NAVEFAHGVGYHOGCCIAASRLFVE-ESIYD 317
QY 304 KFHGVHSGSGGQORWLGONPLEPKRTHGPFVDSKSOYDRVLGNIDVG-KDTAQLLTGVGR 362
DB 318 EF--VRR--SVERAKKVVNLPTPGVSGQPQIDKEQYEKILDLTESGKKEGKACGCGGP 374
QY 363 KGDKGFAIEPTIFVNPKPGSKIMPEEIFGPVLSIKTFTEEAIEANDTTYGLASVIVT 422
DB 375 WGNKGFIQPTVFSVTDMDRIAKEEIFGPVQOIMKFKSLDDVVKRANNTFYGLSAGIFT 434
QY 423 KSLNGLRVSALTEGVSINFPETQTPFGGMKQSGSGRELGEGLKAYLEPTKTI 482
DB 435 NDIKAITVSALQSGTVMVNCYSVVSACQPFPGKMGSGRELGEYGFHEYTEVKTVTI 494
QY 483 HVN 485
DB 495 KIS 497

Search completed: April 27, 2002, 07:48:21
Job time: 6728 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:58:35 ; Search time 24.91 Seconds
(without alignments)
136.751 Million cell updates/sec

Title: US-09-882-694A-5
Perfect score: 2349
Sequence: 1 MVLSPEYKSELFINNEFVS.....EGLKAYLEPKTINIHVNI 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 40831 seqs, 6994805 residues

Total number of hits satisfying chosen parameters: 40831

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: Pending_Patents_AA-New:*
 - 2: /cgn2_6/ptodata/2/paa/PCIT_NEW_COMB.pep.*
 - 3: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
 - 4: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
 - 5: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
 - 6: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
 - 7: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
 - 8: /cgn2_6/ptodata/2/paa/US50_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | % Match | Length | ID | Description |
|------------|-------|-------------|---------|--------|---------------------|--------------------|
| 1 | 561.5 | 22.0 | 528 | 5 | US-09-573-655B-127 | Sequence 127, App |
| 2 | 394 | 15.5 | 607 | 5 | US-09-573-655B-1086 | Sequence 1086, App |
| 3 | 261 | 10.2 | 108 | 6 | US-10-106-698-6298 | Sequence 6298, App |
| 4 | 233.5 | 9.2 | 1185 | 5 | US-09-895-913A-246 | Sequence 246, App |
| 5 | 96.5 | 3.8 | 1292 | 5 | US-09-573-655B-1629 | Sequence 1629, App |
| 6 | 96.5 | 3.8 | 1292 | 5 | US-09-573-655B-2282 | Sequence 2282, App |
| 7 | 94 | 3.7 | 573 | 5 | US-09-847-637B-7 | Sequence 7, Appl1 |
| 8 | 93 | 3.6 | 518 | 5 | US-09-540-209B-6751 | Sequence 6751, App |
| 9 | 91.5 | 3.6 | 612 | 5 | US-09-573-655B-852 | Sequence 852, App |
| 10 | 91.5 | 3.6 | 612 | 5 | US-09-573-655B-1114 | Sequence 1114, App |
| 11 | 90 | 3.5 | 735 | 5 | US-09-978-403A-74 | Sequence 74, Appl |
| 12 | 90 | 3.5 | 735 | 5 | US-09-978-544A-74 | Sequence 74, Appl |
| 13 | 90 | 3.5 | 735 | 5 | US-09-978-681A-74 | Sequence 74, Appl |
| 14 | 90 | 3.5 | 735 | 5 | US-09-978-757A-74 | Sequence 74, Appl |
| 15 | 90 | 3.5 | 735 | 5 | US-09-978-564A-74 | Sequence 74, Appl |
| 16 | 90 | 3.5 | 735 | 5 | US-09-999-831A-74 | Sequence 74, Appl |
| 17 | 90 | 3.5 | 735 | 5 | US-09-999-829A-74 | Sequence 74, Appl |
| 18 | 90 | 3.5 | 735 | 6 | US-10-013-921A-74 | Sequence 74, Appl |
| 19 | 90 | 3.5 | 735 | 6 | US-10-013-929A-74 | Sequence 74, Appl |
| 20 | 90 | 3.5 | 735 | 6 | US-10-013-918A-74 | Sequence 74, Appl |
| 21 | 90 | 3.5 | 735 | 6 | US-10-017-082A-74 | Sequence 74, Appl |
| 22 | 88 | 3.5 | 573 | 5 | US-09-847-637B-8 | Sequence 8, Appl1 |
| 23 | 86 | 3.4 | 529 | 5 | US-09-540-209B-9534 | Sequence 9534, App |
| 24 | 85 | 3.3 | 395 | 5 | US-09-540-209B-9876 | Sequence 9876, App |
| 25 | 85 | 3.3 | 395 | 6 | US-10-108-605-55 | Sequence 55, Appl |
| 26 | 84.5 | 3.3 | 402 | 6 | US-10-036-507-14 | Sequence 14, Appl |

| | | | | | | |
|----|------|-----|------|---|---------------------|-------------------|
| 27 | 80 | 3.1 | 627 | 5 | US-09-540-209B-5747 | Sequence 5747, Ap |
| 28 | 80 | 3.1 | 1207 | 5 | US-09-573-655B-2161 | Sequence 2161, Ap |
| 29 | 80 | 3.1 | 1207 | 5 | US-09-573-655B-2199 | Sequence 2199, Ap |
| 30 | 78.5 | 3.1 | 1326 | 5 | US-09-540-209B-5820 | Sequence 5820, Ap |
| 31 | 76.5 | 3.0 | 1090 | 5 | US-09-573-655B-956 | Sequence 956, App |
| 32 | 76 | 3.0 | 520 | 5 | US-09-540-209B-9731 | Sequence 9731, Ap |
| 33 | 75.5 | 3.0 | 519 | 5 | US-09-540-209B-8693 | Sequence 8693, Ap |
| 34 | 75.5 | 3.0 | 748 | 5 | US-09-573-655B-1010 | Sequence 1010, Ap |
| 35 | 75 | 2.9 | 1138 | 6 | US-10-036-328A-6 | Sequence 6, Appl1 |
| 36 | 75 | 2.9 | 1210 | 6 | US-10-036-328A-2 | Sequence 2, Appl1 |
| 37 | 74.5 | 2.9 | 820 | 5 | US-09-540-209B-6041 | Sequence 6041, Ap |
| 38 | 74.5 | 2.9 | 937 | 5 | US-09-540-209B-7544 | Sequence 7544, Ap |
| 39 | 74.5 | 2.9 | 1068 | 6 | US-10-109-310-7 | Sequence 7, Appl1 |
| 40 | 74.5 | 2.9 | 1092 | 1 | PCT-US01-45053-18 | Sequence 18, Appl |
| 41 | 73.5 | 2.9 | 382 | 5 | US-09-540-209B-5992 | Sequence 5992, Ap |
| 42 | 73.5 | 2.9 | 409 | 5 | US-09-540-209B-9290 | Sequence 9290, Ap |
| 43 | 73.5 | 2.9 | 1021 | 5 | US-09-540-209B-7716 | Sequence 7716, Ap |
| 44 | 73.5 | 2.9 | 1166 | 5 | US-09-972-115A-6 | Sequence 6, Appl1 |
| 45 | 73.5 | 2.9 | 1450 | 5 | US-09-972-211-67 | Sequence 67, Appl |

ALIGNMENTS

RESULT 1
US-09-573-655B-127
; Sequence 127, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVVEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptide
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573.655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 127
; LENGTH: 528
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-127

Query Match 22.0%; Score 561.5; DB 5; Length 528;
Best Local Similarity 32.2%; Pred. No. 6.1e-43;
Matches 160; Conservative 82; Mismatches 202; Indels 53; Gaps 15;

| | | |
|----|-----|---|
| QY | 9 | KSELFNNFEVSKGSRRLTLPNWDSEIVATDVHANAADVDVAASVQAVKGPWK 68 |
| DB | 52 | RTQGLIGGKWLDSYDNKTIKVNPNPATGELIA-DVACMTKETNDIAISSYFATS--WSR 108 |
| QY | 69 | FTGAQRAACMLKFADIAEKNAEKARLESPTGRFVPMITHFDIPNMVSVFVYAGWADK 128 |
| DB | 109 | LTAGERSKVLRRWYDLIIAHKEELGLITLEQKPKLEAIG-EVAYGASFLEYAEEAKR 167 |
| QY | 129 | IAGKTPEDENGPNWR-----YEPMGVCAGIASNNATFLYVGWKIAPALAACGSI 181 |
| DB | 168 | VYGIIP-----PNLSDRLLVLKQFVGVGVAITPWNPLAMITKVGKVPALASCTV 222 |
| QY | 182 | ASEKSLPGLVLGAPFAEAGPPGVVQFLTG-ARVTGEALASHMDIAKISTRSVGGGRA 240 |
| DB | 223 | PSELTPLTALAAALQAGVPPGALNVVMGNAPEIGDALLTSPOVKRITFTGTAVGKK 282 |
| QY | 241 | VKOATLKSNNKVTLEL-GEKPTIVFNEAPLE-RQSGESAKDFSGFGIWPVPSCLLV 298 |
| DB | 283 | L-MAAAPTVKVKVSLGELGNAPSIIVDDADLDVAVKGLAAKFRNSGOTVCANRLVQ 341 |
| QY | 299 | G---NLAKEF-----HGVRHSGGGGQRLWGQNPLEPKTHGTFVYDKSDYRV 345 |
| DB | 342 | GIYDKFAEAFSEAVOKLEVGDGFRDGT-----TQGPLINDAAVQVKET 384 |
| QY | 346 | NI-DVCKDQATQLLTGVRGDKGKGFAL-EPTIVNPKPGSKIWFEIFGPVLISIKTF 403 |

Db 385 FVQDAVSKAKIIG-GKRHSLGHTYEPYIVIRDSNMIMSKKEIFGPVAPLIRFKTEE 443
Qy 404 EAEIANDTTYGLASVIYTKSLNGLRVSALETGCVSNFPIPETQPPFGMKQSSG 463
Db 444 DAIRIANDTTAGLAAYITNSVORSVRVFEALRYGLVGVNEGLISTEVAPEGGVQSGIG 503
Qy 464 RELGEEGLKAYLEPKTI 480
Db 504 REGSKYGMDEYLEIKYV 520

RESULT 2
US-09-573-655B-1086
; Sequence 1086, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573,655B
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1086
; LENGTH: 607
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-1086

Query Match 15.5%; Score 394; DB 5; Length 607;
Best Local Similarity 29.6%; Pred. No. 7.7e-28;
Matches 144; Conservative 64; Mismatches 24; Indels 38; Gaps 15;

Qy 14 INNEFVSKGSERLTLNPNWDESTVATDVHANAADVSAVAASQAVKKGPKKFTGQA 73
Db 118 IGGSFVQSQSSFDVINPATQE-VYSKVPILTNEEFKAASAAKQAPPL--WRWTPITT 174
Qy 74 RAACMLKPADIAEKNAEKLALESPTGRVSMITHFDIPNMVSVFRYAGWADKIAGT 133
Db 175 RQVMLKFOELIRKNMCKLAMNITTEQGGTKL-DSHGDIIFRGLIEVVEHACGMATLQMG 233
Qy 134 FPE-DNGKPNWRY-EPMGVCAGIASWNATFLYVGVKIAFALAAAGCSFIFKASEKSPGL 191
Db 234 LPNYSNGVDIYSIREPLGVGAGICPFNPAMIPLMFPVAVTCGNTFILKPEKDPGASV 293
Qy 192 GLAPLPAEAGFPQGVVQFLTGARTGEALASHMDIAKISFTSVSGGAVKQATLKSMM 251
Db 294 ILAELAMEAGLPDGVNLVHVTGNTVNAICDDEDIRAVSVGSNTAGMHI-YARAAAGK 352
Qy 252 RVTLELGEK-PTIVFNEAPLERO-SGESAKDFSKFGQIWWPPSCLLV-----QW-GNLAE 303
Db 353 RIQSNMGAKNHLGLVLPDANIDATLALLAAGFCAAGQRCMALSTVVFYGDAKSWEDKLVE 412
Qy 304 KPHG--VRHGSFGCCQWLQCNPLEPKRTHGPPVDKSDYDRVLGNIDVGDKTAQLITGV 361
Db 413 RAKALKVTCGS-----EPDADLGPVVISQAKERICRLIQSGVDGAKLLLDG 459
Qy 362 RK----GDKGFALEPTIFVNPKPGSKINFEIFGPVLSIKTFKTEEEAIEIANDTTYGL 416
Db 460 RDIIVPGVYKGNFIPILSGVTPDMCYKEEIFGPVLVCMQANSFDEAISINKNKYGN 519
Qy 417 ASVIYTKSLNGLRVSALETGCVSNFPP-IPETQTPFGMKQSSG--RELGEGLKA 473
Db 520 GNAIFTSGGAARKFQMDIEAGQIGINVPILPPLPFPSFTGNKASFAGDLNFPYKAGVDF 579
Qy 474 YLEPKTI 480
Db 580 FTQIKTV 586

RESULT 3
US-10-106-698-6298

; Sequence 6298, Application US/10106698
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypep
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6298
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-6298

Query Match 10.2%; Score 261; DB 6; Length 108;
Best Local Similarity 52.4%; Pred. No. 5.7e-17;
Matches 54; Conservative 18; Mismatches 31; Indels 0; Gaps 0;

Qy 382 SKIWFEITGPVLSIKTFKTEEEAIEIANDTTYGLASVIYTKSLNGLRVSALETGCVS 441
Db 1 AKTAEIEFGPVNQILKFKTIEEVGRANNSTYGLAAVFTKDLKANYLSQALQAGTVW 60
Qy 442 INFFPIPETQTPFGMKQSGRELGEGLKAYLEPKTINIHV 484
Db 61 VNCYDVFGAQSPGPGYKMGSGRELGEYGLQAYTEVKTVTVKV 103

RESULT 4
US-09-895-913A-246
; Sequence 246, Application US/09895913A
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding Novel Helicobacter Polypeptides in the Helicobacter
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895,913A
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/881,227
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 246
; LENGTH: 1185
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-895-913A-246

Query Match 9.2%; Score 233.5; DB 5; Length 1185;
Best Local Similarity 26.1%; Pred. No. 4.8e-13;
Matches 117; Conservative 64; Mismatches 214; Indels 53; Gaps 18;

Qy 42 VHVANAADVDSAAVSQAVKKGPKKFTGAGRAACMLKFAADLAEKNAEKLALESPTG 101
Db 538 VHLADKEAILKALEVAKSDKRSFSQKSET--EIHALMSQTAQLFRERRGDLIGSALEVG 595
Qy 102 RPYSMITHFDIPNMVSVFRYAGWADKIAGKTFPEDNCKPNRYEPMCVACAGIASWN-AT 160
Db 596 KTFAE-TDAEVSEADLFLEY-----PYSLRVLOEQNTKT--QFTPKGVGVVIAPWNPV 647

QY 161 FLYGVWKTAPALACGSFIFKASEKSPGLVGLAPLFAEAGPPGVVQPLTGARVTEAL 220
Db 648 GISVG-TIAPLARGNRIYKPSLSVTGYKICECFWADAGVPRDALIYLPs-----KGSdI 703
QY 221 ASH-MDIAKTSFTSVSGGGRVAVKQATLKSNNKRVTLLEL-----GEKPTTFVNEAPLEROS 274
Db 704 SEHLRDESIOFA-ILTGGEDTAYKMLKAN---PTLALSNETGKKNATIVSKMADRd--- 756
QY 275 GESAKD-----FSFEGQIWWPPSCLLVQWGNLAEK-----FHGVHRHSGGCGQRMWGN 323
Db 757 -OAIKNVIHSAFSSGQKCSATSLVLEKEVEYEDENFKKTLIDATLSLVG-----D 807
QY 324 PLEPKRTHGPPVDKSOQDRVLGNIDVGKOTQAQLLTGVGRKDKGFAIEPTIFVNPKPQSGK 383
Db 808 PFDEKNGIGALADKPn-EKVIKAIDELKSYENYEIPVSFVNDNPNLKMPSIKYGTKKGF 866
QY 384 IWFEEIFGPVLSIKTFTEEEAEIANDTTYGLASVIYTKSLNRGLRVSSALETGQVSN 443
Db 867 THQTELETPILSVMEAKDLDEAIEANSTGYLTSALDESIDEREWBYLERIPAGNIYN 926
QY 444 PFEIPE--TOTPFMGKOS--GSGRELG 467
Db 927 KPTTGAVLRQPPGVSXKSAVGGRKVG 954

RESULT 5

US-09-573-655B-1629
; Sequence 1629, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVIEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573,655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1629
; LENGTH: 1292
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-1629

Query Match 3.8%; Score 96.5; DB 5; Length 1292;
Best Local Similarity 22.0%; Pred. No. 1;
Matches 118; Conservative 68; Mismatches 201; Indels 149; Gaps 30;

QY 10 SELFIN-----NEFVSSKG-----SERLTITNPWDESTVATDVHVNAA---DVDsAVA 55
Db 746 SGLFYNQIIGNALVSMYKIGEMSESRVLLQMP-RRDVVANNALIGGYAEDEDPDKALA 804
QY 56 ASVQAVKKGPKKFP--TGAQRAACMLKPADLAENKAEKLARLESPTGRPVs---MITHF 110
Db 805 AFQTMREVGSNNYITVVSLSACLLP-GDILLER-----GKPLHAYIVSAGF 850
QY 111 DIPNNV--SVFRYAGWADKIAGTFFPDNGKPNRWPEPMGVCGAGIASWATFLYVGWKI 168
Db 851 ESDEHVKNLSITWAKCGDLSSQDL--FNGLDN-----RNIIITWNA----- 890
QY 169 APALAAGCSFIFKASEKSPGLVGLAPLFAEAGPPGVVQPLTGARVTEALASHMDIAK 228
Db 891 --MLAAN-----AHHGHGEVLLKLVSKMRSGVSLD----- 919
QY 229 ISFTSVSGGGRVAVKQATLK--SNMKRVTLLELG-EKPTIVNEAPLERQSGESAKDFSKEG 285
Db 920 -QFSFSEGLSAAKLAULVEEQQLHGLAVLKGFEDHSFIFNAA-----ADMTSKCG 969
QY 286 QT-----WVPPSC--LLVQWGNLAEKPHGVHRHSGFGG-CQRW-----LGQNP-----L 325
Db 970 EIGEVVKKMLPPSVNRLSPSNILISALG--RHGYFEVCAATFHEMLEMGIKPGHVTFSVL 1027
QY 326 EPKRTHTGPFVDK--SOYDRVLGNIDVGKOTQAQLLTGVGRKDKGFAIEPTIFVNP---KP 380

Db 1028 LTACHSHGLVDKGLAYDYMIAIDRFGLEPAIEHCICVTDLLGRSGRLAEAEFTFSKMPMKP 1087
QY 381 GSKIWFE-----EIFGPVLSIKTFTEEEAEIANDTTYGLASVIYTKSLNRGLRVSSAL 435
Db 1088 NDLVWRSLLASCKIHGNLDRGRKAAENLSKLEPEDDSVYVLSNNMFA-TTGRWEDVENVR 1146
QY 436 ETGGVSVNFPPIPETOTPFMGKOSGSGRELGE---GLKAYLEPKTINIHNIE 487
Db 1147 KQMG-----FKNIKK-----KOACSWVKLDRKVSFGIGDRTHPQTMEIYAKLE 1190
RESULT 6
US-09-573-655B-2282
; Sequence 2282, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVIEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573,655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2282
; LENGTH: 1292
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-2282

Query Match 3.8%; Score 96.5; DB 5; Length 1292;
Best Local Similarity 22.0%; Pred. No. 1;
Matches 118; Conservative 68; Mismatches 201; Indels 149; Gaps 30;

QY 10 SELFIN-----NEFVSSKG-----SERLTITNPWDESTVATDVHVNAA---DVDsAVA 55
Db 746 SGLFYNQIIGNALVSMYKIGEMSESRVLLQMP-RRDVVANNALIGGYAEDEDPDKALA 804
QY 56 ASVQAVKKGPKKFP--TGAQRAACMLKPADLAENKAEKLARLESPTGRPVs---MITHF 110
Db 805 AFQTMREVGSNNYITVVSLSACLLP-GDILLER-----GKPLHAYIVSAGF 850
QY 111 DIPNNV--SVFRYAGWADKIAGTFFPDNGKPNRWPEPMGVCGAGIASWATFLYVGWKI 168
Db 851 ESDEHVKNLSITWAKCGDLSSQDL--FNGLDN-----RNIIITWNA----- 890
QY 169 APALAAGCSFIFKASEKSPGLVGLAPLFAEAGPPGVVQPLTGARVTEALASHMDIAK 228
Db 891 --MLAAN-----AHHGHGEVLLKLVSKMRSGVSLD----- 919
QY 229 ISFTSVSGGGRVAVKQATLK--SNMKRVTLLELG-EKPTIVNEAPLERQSGESAKDFSKEG 285
Db 920 -QFSFSEGLSAAKLAULVEEQQLHGLAVLKGFEDHSFIFNAA-----ADMTSKCG 969
QY 286 QT-----WVPPSC--LLVQWGNLAEKPHGVHRHSGFGG-CQRW-----LGQNP-----L 325
Db 970 EIGEVVKKMLPPSVNRLSPSNILISALG--RHGYFEVCAATFHEMLEMGIKPGHVTFSVL 1027
QY 326 EPKRTHTGPFVDK--SOYDRVLGNIDVGKOTQAQLLTGVGRKDKGFAIEPTIFVNP---KP 380
Db 1028 LTACHSHGLVDKGLAYDYMIAIDRFGLEPAIEHCICVTDLLGRSGRLAEAEFTFSKMPMKP 1087
QY 381 GSKIWFE-----EIFGPVLSIKTFTEEEAEIANDTTYGLASVIYTKSLNRGLRVSSAL 435
Db 1088 NDLVWRSLLASCKIHGNLDRGRKAAENLSKLEPEDDSVYVLSNNMFA-TTGRWEDVENVR 1146
QY 436 ETGGVSVNFPPIPETOTPFMGKOSGSGRELGE---GLKAYLEPKTINIHNIE 487
Db 1147 KQMG-----FKNIKK-----KOACSWVKLDRKVSFGIGDRTHPQTMEIYAKLE 1190
RESULT 7

US-09-847-637B-7

```

: Sequence 7, Application US/09847637B
: GENERAL INFORMATION:
: APPLICANT: Naparstek, Yakov
: APPLICANT: Umanasky, Rina
: APPLICANT: Kashi, Yechezkel
: TITLE OF INVENTION: NOVEL AMINO ACID SEQUENCES AND THE D
: TITLE OF INVENTION: THE AMINO ACID SEQUENCES AND THE D
: TITLE OF INVENTION: SEQUENCES AND THE D
: FILE REFERENCE: 13125-002001
: CURRENT APPLICATION NUMBER: US/09/847,637B
: CURRENT FILING DATE: 2001-05-02
: PRIOR APPLICATION NUMBER: PCT/IL99/00595
: PRIOR FILING DATE: 1999-11-04
: PRIOR APPLICATION NUMBER: 60/107,213
: PRIOR FILING DATE: 1998-11-05
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 7
: LENGTH: 573
: TYPE: PRT
: ORGANISM: Rattus norvegicus
: US-09-847-637B-7

```

Query Match 3.7%; Score 94; DB 5; Length 573;
Best Local Similarity 24.9%; Pred. NO. 0.56;
Matches 53; Conservative 22; Mismatches 70; Indels 68; Gaps 10;

RESULT.T

```

US-09-540-209B-6751
; Sequence 6751, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Bretton
; TITLE OF INVENTION: NUCLEIC ACID AND A
; TITLE OF INVENTION: FOR DIAGNOSTICS A
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 6751
; LENGTH: 518
; TYPE: prt
; ORGANISM: B.fragilis
US-09-540-209B-6751

```

Query Match 3.6%; Score 93; DB 5; Length 518;
Best Local Similarity 24.9%; Pred. No. 0.59;
Matches 63; Conservative 20; Mismatches 86; Indels 84; Gaps 14;

| | | | | | |
|----|-----|---|---------------------------------|---------------------------|----------|
| Db | 179 | GEYDPGNGDIAKNA | ---QNGFP-WR----- | SGOAIW----- | 2099 |
| Qy | 179 | IFKASEKSPGLVGLAPLF | ----- | APAGEF-PPGVQVFLTGARV----- | TGEA 219 |
| Db | 210 | DHGSIAGSALCKMGIIDYFRIPKRWYRYNAVYKGTTPPEWPOEGTPARISLVADRTDNI | 2699 | | |
| Qy | 220 | LASHMDIAKTSFTSRVSGGGRVAKQATLKSNNKRVTEL | ---GEKPT--- | IVF-NEAPLE 271 | |
| Db | 270 | KADGTDDVMUISITILDANGRPV----- | SNSPAVKDILSGPGEFTGTISILFEKESDIR | 323 | |
| Qy | 272 | ROSGESAKDFSKF | 284 | | |
| Db | 324 | ILDGKAAIEFRSY | 336 | | |

RESULT

```

US-09-573-655B-952
:
: GENERAL INFORMATION:
:
: APPLICANT: SOLOVYEV, Victor and TROUKH
:
: TITLE OF INVENTION: Sequence-Determin
:
: TITLE OF INVENTION: Thereby
:
: FILE REFERENCE: 2750-0876P
:
: CURRENT APPLICATION NUMBER: US/09/573
:
: CURRENT FILING DATE: 2000-05-18
:
: NUMBER OF SEQ ID NOS: 3281
:
: SOFTWARE: PatentIn version 3.0
:
: SEQ ID NO 952
:
: LENGTH: 612
:
: TYPE: prt
:
: ORGANISM: Arabidopsis thaliana
US-09-573-655B-952

```

Query Match 3.68; Score 91.5; DB 5; Length 612;

RESULT 10

```

US-09-573-655B-1114
; Sequence 1114, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVIEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-0876P
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1114
; LENGTH: 612
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-1114

Query Match          3.6%; Score 91.5; DB 5; Length 612;
Best Local Similarity 19.7%; Pred. No. 1;
Matches      88; Conservative    70; Mismatches 168; Indels 121; Gaps   20;

QY      4 SPEYKSELFINNEFSSGSRSLRLTNPWDESVATDVHANAADVDASAASVOAQ-K 62
Db      180 SDSDNVNKSVEDEGEDERDVEDIVPEVERIDEI-----ANVDMEEAMAKQLCM 234
               : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      63 KGPWKFTGAQRAA--CMLK-----FADLAEKNAEKLARLESLTGRPSVKMTH---FDI 112
Db      235 YFPASYTYTKMKLATRCYISEVLKTFADLE-----HPLTNVENNYPMEH 278
               : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     113 PNKVSVFRYYAGWAKIAG-----KTPEONGKPNMRYEPMGVCAGIASWNATFLYVGW 166
Db     279 PSFKHIYHLPSGYTHKLRMMWLFLRTYSIEKKKEVW-YFKTGv----- 321
               : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     167 KIAPALAAAGCSIFRASEKSPGLVGLAPLFAEGAFPFGVVQFLTGARV----TGEALAS 222
Db     322 -----IRREDVKTLK--MEMEPAKSKORLMNAVLYFLTSSIVPTKTGER-AS 366
               : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     223 HMD-----IAKISFTRSGGVRQVQKATLKSNMKRVTLERGEPTTVFNEAPLERQSQE 276
Db     367 PIDDFCVRAASDLTFCKTPFMWGRISEFYMLKLSISHTLDHENGVVVPNI---QSP----- 416
               : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     277 SAKDFSKFGQIW-VPPSCLLLVQ-----GNLAKEFHGVRSFGCGCORWLGQNPLEPKR 329
Db     417 -----WPVPGFCVPLEFAFEAFIPSLRERFIEXEGAHAGCPCKMYN---FKR 462
               : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     330 THGPFYDKSQYDRVLGNIDV-----GKDTAQI--ITGVGRKGDKGFAGEIPTIFYNPK 379
Db     463 TEMKRFTLQINHVLGTTEVIESITREKAEEVPLIAETIGVEDDDKHVDHVDSMMKRLG 522
               : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     380 PGSKIWFEEELGFPVLSIKTFKTEREAI 406
Db     523 QRREIRFEVINEDVHARKAPNEEEV 549
               : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
US-09-978-403A-74
; Sequence 74, Application US/09978403A
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleon
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Cao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher

```

; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080328
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080334
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081071
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081203
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081229
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081955
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081817
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081819
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081952
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082804
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082700
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082796
; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: 60/083336
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083392
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083545
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083554
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083558
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083500
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083742
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06

; PRIOR APPLICATION NUMBER: 60/084441
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084637
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084598
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084627
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085339
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085338
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085689
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697
; PRIOR FILING DATE: 1998-05-15

Query Match 3.5%; Score 90; DB 5; Length 735;

Best Local Similarity 21.3%; Pred. No. 1.8;

Matches 63; Conservative 32; Mismatches 75; Indels 126; Gaps 14;

Qy 239 RAVKQATLKSNMKRVTLLEGEKPTIVFNEAPLERQSESAKDFSGQIWWPPSCLLVQW 298
Db 200 RRKRETLKAT-KYVEL-----VIVADNREFORQ-----GKDEKVKO-----RUIEI 241
Qy 299 GNLAKEPH---GVRHGSFGGQQRW-----LGQNP-----LEPKRTHGPF 334
Db 242 ANHVDKIFYRPLNIRI-VLVGVEVWMDMKCSVSQDPFTSLHFLDWRKMKLLPRKSH--- 297
Qy 335 VDKSQYDRVLGNIDVGKDTAQLLTGVGKRGDK-GFAIEPTIFVNPKPGSKIWFEEIGPV 393
Db 298 -----DNAQLVSGVYFGTTTGM-----PI 318
Qy 394 LSIKTFKTEEEATEIANDTTYGLASVIYTK-----SLNRLRVSSALETGGVSIN 443
Db 319 MSMTADQSGGIYMDHSDNPLGRAVTLAHELGHNFHGHNDTLDHGSCCMAVEKGGCINM 378
Qy 444 -----FPP-----IPETQTFPGCMKQSGSGRELEGE 469
Db 379 ASTGYPPPMVFSSCSRRKDLTSLKGMGVCLFNLPLPEYRESFGQKCGNRFEVEGEE 434

RESULT 12

US-09-978-544A-74

; Sequence 74, Application US/09978544A

; GENERAL INFORMATION:

; APPLICANT: Ashkenazi, Avi

; APPLICANT: Baker Kevin P.

; APPLICANT: Botstein, David

; APPLICANT: Desnoyers, Luc

; APPLICANT: Eaton, Dan

APPLICANT: Ferrara, Napoleon
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gertitsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Shelton, David L.
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2630PIC13
CURRENT APPLICATION NUMBER: US/09/978,544A
CURRENT FILING DATE: 2002-03-19
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064249
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077641
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078004
PRIOR FILING DATE: 1998-03-13
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079920
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/079923
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/080105
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/083336
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083554
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083558
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559

```

; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083500
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083742
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084441
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084637
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084598
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084627
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085339
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085338
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085689
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697
; PRIOR FILING DATE: 1998-05-15

Query Match          3.5%; Score 90; DB 5; Length 735;
Best Local Similarity 21.3%; Pred. No. 1.8;
Matches 63; Conservative 32; Mismatches 75; Indels 126; Gaps 14;

QY 239 RAVKQATLSNMKRVLTLEGEKPTIVFNAPLERQSGESAKDFSGQIWPVPPSCLLVQM 298
DB 200 RRHKRETLKAT-KYVEL-----VIVADNREFQK-----GKDLKVKO-----RLIEI 241
QY 299 GNLAERFH---GVRHGSFGGQKRW-----LCQNP-----LEPKRTHGPF 334
DB 242 ANHVDKFRPLNIRI-VLVGVEVWMDKCSVSQDPFTSLHEFLDWRKMKLLPRKSH--- 297
QY 335 VDKSQDVRVLGNTDVKCKDAQLLTGVGRKGDK-GFAIEFTIPVNPKPSKIWFELIFGPV 393
DB 298 -----DNAQLVSGYVFOGTTIGMA-----PI 318
QY 394 LSIKTFKEEAEIETANDTYGLASVIYTK-----SLNRGLRVSSALETGCVSIN 443
DB 319 MSMTAQDSQGIWMDHSDNPLGAAVTLAHELGHNFNMHDTLDRCSCQMAVEKGGCIWN 378
QY 444 ----PFP-----IPETQTPFGGMKQSGSGRELGE 469
DB 379 ASTGYFPFPMVFSKRSKDKLSTSLKGMGVCLNLFVRESFGKCGNRFVEEGE 434
```

```

RESULT 13
US-09-978-681A-74
; Sequence 74, Application US/09978681A
; GENERAL INFORMATION:
; APPLICANT: ASHENAZI, Avi
; APPLICANT: Baker Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleon
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Oiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kijavin, Ivar J.
; APPLICANT: Kuo, Sophia S.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Shelton, David L.
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2630P1C18
; CURRENT APPLICATION NUMBER: US/09/978,681A
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: 09/918585
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/064249
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077641
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077791
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/078004
; PRIOR FILING DATE: 1998-03-13
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078936
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079689
```


| | | |
|----|-------------------------------------|--|
| 1 | PRIOR FILING DATE: 1998-03-27 | |
| 2 | PRIOR APPLICATION NUMBER: 60/079663 | |
| 3 | PRIOR FILING DATE: 1998-03-27 | |
| 4 | PRIOR APPLICATION NUMBER: 60/079728 | |
| 5 | PRIOR FILING DATE: 1998-03-27 | |
| 6 | PRIOR APPLICATION NUMBER: 60/079786 | |
| 7 | PRIOR FILING DATE: 1998-03-27 | |
| 8 | PRIOR APPLICATION NUMBER: 60/079920 | |
| 9 | PRIOR FILING DATE: 1998-03-30 | |
| 10 | PRIOR APPLICATION NUMBER: 60/079923 | |
| 11 | PRIOR FILING DATE: 1998-03-30 | |
| 12 | PRIOR APPLICATION NUMBER: 60/080105 | |
| 13 | PRIOR FILING DATE: 1998-03-31 | |
| 14 | PRIOR APPLICATION NUMBER: 60/080107 | |
| 15 | PRIOR FILING DATE: 1998-03-31 | |
| 16 | PRIOR APPLICATION NUMBER: 60/080165 | |
| 17 | PRIOR FILING DATE: 1998-03-31 | |
| 18 | PRIOR APPLICATION NUMBER: 60/080194 | |
| 19 | PRIOR FILING DATE: 1998-03-31 | |
| 20 | PRIOR APPLICATION NUMBER: 60/080327 | |
| 21 | PRIOR FILING DATE: 1998-04-01 | |
| 22 | PRIOR APPLICATION NUMBER: 60/080328 | |
| 23 | PRIOR FILING DATE: 1998-04-01 | |
| 24 | PRIOR APPLICATION NUMBER: 60/080333 | |
| 25 | PRIOR FILING DATE: 1998-04-01 | |
| 26 | PRIOR APPLICATION NUMBER: 60/080334 | |
| 27 | PRIOR FILING DATE: 1998-04-01 | |
| 28 | PRIOR APPLICATION NUMBER: 60/081070 | |
| 29 | PRIOR FILING DATE: 1998-04-08 | |
| 30 | PRIOR APPLICATION NUMBER: 60/081049 | |
| 31 | PRIOR FILING DATE: 1998-04-08 | |
| 32 | PRIOR APPLICATION NUMBER: 60/081071 | |
| 33 | PRIOR FILING DATE: 1998-04-08 | |
| 34 | PRIOR APPLICATION NUMBER: 60/081195 | |
| 35 | PRIOR FILING DATE: 1998-04-08 | |
| 36 | PRIOR APPLICATION NUMBER: 60/081203 | |
| 37 | PRIOR FILING DATE: 1998-04-09 | |
| 38 | PRIOR APPLICATION NUMBER: 60/081229 | |
| 39 | PRIOR FILING DATE: 1998-04-09 | |
| 40 | PRIOR APPLICATION NUMBER: 60/081955 | |
| 41 | PRIOR FILING DATE: 1998-04-15 | |
| 42 | PRIOR APPLICATION NUMBER: 60/081817 | |
| 43 | PRIOR FILING DATE: 1998-04-15 | |
| 44 | PRIOR APPLICATION NUMBER: 60/081819 | |
| 45 | PRIOR FILING DATE: 1998-04-15 | |
| 46 | PRIOR APPLICATION NUMBER: 60/081952 | |
| 47 | PRIOR FILING DATE: 1998-04-15 | |
| 48 | PRIOR APPLICATION NUMBER: 60/081838 | |
| 49 | PRIOR FILING DATE: 1998-04-15 | |
| 50 | PRIOR APPLICATION NUMBER: 60/082568 | |
| 51 | PRIOR FILING DATE: 1998-04-21 | |
| 52 | PRIOR APPLICATION NUMBER: 60/082569 | |
| 53 | PRIOR FILING DATE: 1998-04-21 | |
| 54 | PRIOR APPLICATION NUMBER: 60/082704 | |
| 55 | PRIOR FILING DATE: 1998-04-22 | |
| 56 | PRIOR APPLICATION NUMBER: 60/082804 | |
| 57 | PRIOR FILING DATE: 1998-04-22 | |
| 58 | PRIOR APPLICATION NUMBER: 60/082700 | |
| 59 | PRIOR FILING DATE: 1998-04-22 | |
| 60 | PRIOR APPLICATION NUMBER: 60/082797 | |
| 61 | PRIOR FILING DATE: 1998-04-22 | |
| 62 | PRIOR APPLICATION NUMBER: 60/082796 | |
| 63 | PRIOR FILING DATE: 1998-04-23 | |
| 64 | PRIOR APPLICATION NUMBER: 60/083336 | |
| 65 | PRIOR FILING DATE: 1998-04-27 | |
| 66 | PRIOR APPLICATION NUMBER: 60/083322 | |
| 67 | PRIOR FILING DATE: 1998-04-28 | |
| 68 | PRIOR APPLICATION NUMBER: 60/083392 | |
| 69 | PRIOR FILING DATE: 1998-04-29 | |
| 70 | PRIOR APPLICATION NUMBER: 60/083495 | |
| 71 | PRIOR FILING DATE: 1998-04-29 | |
| 72 | PRIOR APPLICATION NUMBER: 60/083496 | |
| 73 | PRIOR FILING DATE: 1998-04-29 | |

| | |
|----|-------------------------------------|
| 1 | PRIOR APPLICATION NUMBER: 60/083499 |
| 2 | PRIOR FILING DATE: 1998-04-29 |
| 3 | PRIOR APPLICATION NUMBER: 60/083545 |
| 4 | PRIOR FILING DATE: 1998-04-29 |
| 5 | PRIOR APPLICATION NUMBER: 60/083554 |
| 6 | PRIOR FILING DATE: 1998-04-29 |
| 7 | PRIOR APPLICATION NUMBER: 60/083558 |
| 8 | PRIOR FILING DATE: 1998-04-29 |
| 9 | PRIOR APPLICATION NUMBER: 60/083559 |
| 10 | PRIOR FILING DATE: 1998-04-29 |
| 11 | PRIOR APPLICATION NUMBER: 60/083500 |
| 12 | PRIOR FILING DATE: 1998-04-29 |
| 13 | PRIOR APPLICATION NUMBER: 60/083742 |
| 14 | PRIOR FILING DATE: 1998-04-30 |
| 15 | PRIOR APPLICATION NUMBER: 60/084366 |
| 16 | PRIOR FILING DATE: 1998-05-05 |
| 17 | PRIOR APPLICATION NUMBER: 60/084414 |
| 18 | PRIOR FILING DATE: 1998-05-06 |
| 19 | PRIOR APPLICATION NUMBER: 60/084441 |
| 20 | PRIOR FILING DATE: 1998-05-06 |
| 21 | PRIOR APPLICATION NUMBER: 60/084637 |
| 22 | PRIOR FILING DATE: 1998-05-07 |
| 23 | PRIOR APPLICATION NUMBER: 60/084639 |
| 24 | PRIOR FILING DATE: 1998-05-07 |
| 25 | PRIOR APPLICATION NUMBER: 60/084640 |
| 26 | PRIOR FILING DATE: 1998-05-07 |
| 27 | PRIOR APPLICATION NUMBER: 60/084598 |
| 28 | PRIOR FILING DATE: 1998-05-07 |
| 29 | PRIOR APPLICATION NUMBER: 60/084600 |
| 30 | PRIOR FILING DATE: 1998-05-07 |
| 31 | PRIOR APPLICATION NUMBER: 60/084627 |
| 32 | PRIOR FILING DATE: 1998-05-07 |
| 33 | PRIOR APPLICATION NUMBER: 60/084643 |
| 34 | PRIOR FILING DATE: 1998-05-07 |
| 35 | PRIOR APPLICATION NUMBER: 60/085339 |
| 36 | PRIOR FILING DATE: 1998-05-13 |
| 37 | PRIOR APPLICATION NUMBER: 60/085338 |
| 38 | PRIOR FILING DATE: 1998-05-13 |
| 39 | PRIOR APPLICATION NUMBER: 60/085323 |
| 40 | PRIOR FILING DATE: 1998-05-13 |
| 41 | PRIOR APPLICATION NUMBER: 60/085582 |
| 42 | PRIOR FILING DATE: 1998-05-15 |
| 43 | PRIOR APPLICATION NUMBER: 60/085700 |
| 44 | PRIOR FILING DATE: 1998-05-15 |
| 45 | PRIOR APPLICATION NUMBER: 60/085689 |
| 46 | PRIOR FILING DATE: 1998-05-15 |
| 47 | PRIOR APPLICATION NUMBER: 60/085759 |
| 48 | PRIOR FILING DATE: 1998-05-15 |
| 49 | PRIOR APPLICATION NUMBER: 60/085580 |
| 50 | PRIOR FILING DATE: 1998-05-15 |
| 51 | PRIOR APPLICATION NUMBER: 60/085573 |
| 52 | PRIOR FILING DATE: 1998-05-15 |
| 53 | PRIOR APPLICATION NUMBER: 60/085704 |
| 54 | PRIOR FILING DATE: 1998-05-15 |
| 55 | PRIOR APPLICATION NUMBER: 60/085697 |
| 56 | PRIOR FILING DATE: 1998-05-15 |

Query Match 3.5%; Score 90; DB 5; Length 735;

| | |
|-----------------------|-----------------------------|
| Query match | 3.58; score 50; length 150; |
| Best Local Similarity | 21.38; Pred. No. 1.8; |

| | | | | | |
|-----------------------|------------------|----------------|------------|-----------|-----|
| Best local similarity | 21.89; Accurate | 16; | | | |
| Matches | 63; Conservative | 32; Mismatches | 75; Indels | 126; Gaps | 14; |

QV 239 RAVKOATLKSNNMKRVTLLEGEKPTIVFNEAPLERQSGESAKDESKFGQIWVPPSCLLVQW 298

| | |
|-----|---|
| 200 | RRHRETLKAT-KYVEL-----VIVADNRFRQRQ-----GKDLEKVKQ-----RLIEI 241 |
| Db | |

QY 299 GNLAKEFH--GVRHGSFGCQRW-----LQNP-----LEPKRTHGPF 334

```

_ _ _ _ _ : _ _ _ _ _ : _ _ _ _ _ : _ _ _ _ _ : _ _ _ _ _
242 ANHYDKFYRPLNTRI-VLVGVEVWMDKCSQDPFTSLHEFLDWRKMKLLPRKSH---297
Db

```

QY 335 VDKSQYDRVLGNIDVGKDTAQLLTGYGRKGDK-GFAIEPTIFVNPAPGSKIWFEEIFGPV 393

Db 298 -----DNAQLVSGYFQGTITGMA-----PI 318

;; PRIOR FILING DATE: 1998-04-27
;; PRIOR APPLICATION NUMBER: 60/083322
;; PRIOR FILING DATE: 1998-04-28
;; PRIOR APPLICATION NUMBER: 60/083392
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083495
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083496
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083499
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083545
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083554
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083558
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083559
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083500
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083742
;; PRIOR FILING DATE: 1998-04-30
;; PRIOR APPLICATION NUMBER: 60/084366
;; PRIOR FILING DATE: 1998-05-05
;; PRIOR APPLICATION NUMBER: 60/084414
;; PRIOR FILING DATE: 1998-05-06
;; PRIOR APPLICATION NUMBER: 60/084441
;; PRIOR FILING DATE: 1998-05-06
;; PRIOR APPLICATION NUMBER: 60/084637
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084639
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084640
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084598
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084600
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084627
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084643
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/085339
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085338
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085323
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085582
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085700
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085689
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085579
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085580
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085573
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085704
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085697
;; PRIOR FILING DATE: 1998-05-15

Query Match 3.5%; Score 90; DB 5; Length 735;

Best Local Similarity 21.3%; Pred. No. 1.8;

Matches 63; Conservative 32; Mismatches 75; Indels 136; Gaps 14;

QY 239 RAVKQATLKSNNKRVTELGEKPTIVFNEAPLERQSGESAKDFSKFGIWPVPPSCLLYOW 298

DB 200 RHKREILKAT-KVEL-----VIVADNEFORQ---GKOLEKVQK-----RLIEI 241

QY 299 GNLAERFH--GVRHGSFGGCGW-----LGQNP-----LEPKRTHGPF 334
DB 242 ANIVDFEYRLNIRI-VLVGVVEVNDMDKCSVSQDPFTSLHEFLDWRKMKLLPRKSH--- 297
QY 335 VKSQDYRVLGNIDVGKDTAQLLTGVGRKGDK-GFAIEPTIFVNPKPGSKTWFEIEGVPV 393
DB 298 -----DNAQLVSGVYFGTIGMA-----PI 318
QY 394 LSIKTEKTEEEATEIANDTYGLASVIVTK-----SINRGLRVSSALETGGVSVIN 443
DB 319 MSCTADQSGGIVMDHSDNPICAAVTLAHELGHNFNMHDITLDRCSCQMAVEKGGCIN 378
QY 444 ----FFP-----IPETQTPFGGMMKQSGSGRELGEE 469
DB 379 ASTGYFPMPVFFSSCRKDLTSLEKGMGVCLFNLPEVRESFGQKCGNRFRVEEGEE 434

RESULT 15

US-09-978-564A-74
; Sequence 74, Application US/09978564A
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleon
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Kuo, Sophia S.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoli, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Shelton, David L.
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2630PIC25
; CURRENT APPLICATION NUMBER: US/09/978,564A
; CURRENT FILING DATE: 2001-10-16
; PRIOR APPLICATION NUMBER: 09/918585
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/064249
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077641
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077791

| | |
|-------------------------------------|-------------------------------|
| PRIOR APPLICATION NUMBER: 60/082804 | PRIOR FILING DATE: 1998-04-22 |
| PRIOR APPLICATION NUMBER: 60/082700 | PRIOR FILING DATE: 1998-04-22 |
| PRIOR APPLICATION NUMBER: 60/082700 | PRIOR FILING DATE: 1998-04-22 |
| PRIOR APPLICATION NUMBER: 60/082797 | PRIOR FILING DATE: 1998-04-22 |
| PRIOR APPLICATION NUMBER: 60/082796 | PRIOR FILING DATE: 1998-04-22 |
| PRIOR APPLICATION NUMBER: 60/082796 | PRIOR FILING DATE: 1998-04-22 |
| PRIOR APPLICATION NUMBER: 60/083336 | PRIOR FILING DATE: 1998-04-27 |
| PRIOR APPLICATION NUMBER: 60/083322 | PRIOR FILING DATE: 1998-04-28 |
| PRIOR APPLICATION NUMBER: 60/083392 | PRIOR FILING DATE: 1998-04-29 |
| PRIOR APPLICATION NUMBER: 60/083495 | PRIOR FILING DATE: 1998-04-29 |
| PRIOR APPLICATION NUMBER: 60/083496 | PRIOR FILING DATE: 1998-04-29 |
| PRIOR APPLICATION NUMBER: 60/083499 | PRIOR FILING DATE: 1998-04-29 |
| PRIOR APPLICATION NUMBER: 60/083545 | PRIOR FILING DATE: 1998-04-29 |
| PRIOR APPLICATION NUMBER: 60/083554 | PRIOR FILING DATE: 1998-04-29 |
| PRIOR APPLICATION NUMBER: 60/083558 | PRIOR FILING DATE: 1998-04-29 |
| PRIOR APPLICATION NUMBER: 60/083559 | PRIOR FILING DATE: 1998-04-29 |
| PRIOR APPLICATION NUMBER: 60/083500 | PRIOR FILING DATE: 1998-04-29 |
| PRIOR APPLICATION NUMBER: 60/083742 | PRIOR FILING DATE: 1998-04-30 |
| PRIOR APPLICATION NUMBER: 60/084366 | PRIOR FILING DATE: 1998-05-05 |
| PRIOR APPLICATION NUMBER: 60/084414 | PRIOR FILING DATE: 1998-05-06 |
| PRIOR APPLICATION NUMBER: 60/084441 | PRIOR FILING DATE: 1998-05-06 |
| PRIOR APPLICATION NUMBER: 60/084637 | PRIOR FILING DATE: 1998-05-07 |
| PRIOR APPLICATION NUMBER: 60/084639 | PRIOR FILING DATE: 1998-05-07 |
| PRIOR APPLICATION NUMBER: 60/084640 | PRIOR FILING DATE: 1998-05-07 |
| PRIOR APPLICATION NUMBER: 60/084598 | PRIOR FILING DATE: 1998-05-07 |
| PRIOR APPLICATION NUMBER: 60/084600 | PRIOR FILING DATE: 1998-05-07 |
| PRIOR APPLICATION NUMBER: 60/084627 | PRIOR FILING DATE: 1998-05-07 |
| PRIOR APPLICATION NUMBER: 60/084643 | PRIOR FILING DATE: 1998-05-07 |
| PRIOR APPLICATION NUMBER: 60/085339 | PRIOR FILING DATE: 1998-05-13 |
| PRIOR APPLICATION NUMBER: 60/085338 | PRIOR FILING DATE: 1998-05-13 |
| PRIOR APPLICATION NUMBER: 60/085323 | PRIOR FILING DATE: 1998-05-13 |
| PRIOR APPLICATION NUMBER: 60/085582 | PRIOR FILING DATE: 1998-05-15 |
| PRIOR APPLICATION NUMBER: 60/085700 | PRIOR FILING DATE: 1998-05-15 |
| PRIOR APPLICATION NUMBER: 60/085689 | PRIOR FILING DATE: 1998-05-15 |
| PRIOR APPLICATION NUMBER: 60/085579 | PRIOR FILING DATE: 1998-05-15 |
| PRIOR APPLICATION NUMBER: 60/085580 | PRIOR FILING DATE: 1998-05-15 |
| PRIOR APPLICATION NUMBER: 60/085573 | PRIOR FILING DATE: 1998-05-15 |
| PRIOR APPLICATION NUMBER: 60/085704 | PRIOR FILING DATE: 1998-05-15 |
| PRIOR APPLICATION NUMBER: 60/085697 | PRIOR FILING DATE: 1998-05-15 |

; PRIOR FILING DATE: 1998-05-15

Query Match 3.5%; Score 90; DB 5; Length 735;

Query match: 5.3%; Score 50; 5.3%;
Best Local Similarity 21.3%; Pred. No. 1.8;

Best Local Similarity 21.3%, Freq. NO. 1.0;
Matches 63; Conservative 32; Mismatches 75; Indels 126; Gaps 14;

| | | | |
|----|-----|--|-----|
| Qy | 239 | RAVKAOTLKSNNKRVTLLELGEKPTIFVNEAPLPEROGSGEAKDSKTCOIWVPSPCLLQW | 296 |
| Db | 200 | RHRKRETLKAT-KIVEL-----VIVADNREFQRO-----GKDEKYQK-----RLBZ | 241 |
| Qy | 299 | GNLAEKFH-----GYRHSGFGCCQRW-----LGQNP-----LEPKRTHGPF | 334 |
| Db | 242 | ANHYDKFYPRLNIRI-VLVGVEVMDNDKSCVSQPTFTSLHEFLDWKRMKMLPKRH-- | 297 |
| Qy | 335 | VDKSOYDRVLGNDVGKDTAQLLTGYGVRKGDK-GFAIEPTIFVNPKPGSKINFEELFGPV | 393 |
| Db | 298 | -----DNAQLVSYGYFOGTTIGMA-----PI | 318 |
| Qy | 394 | LSIKTFKTEEBEIAIANDTTYTGILASVIYTK-----SLNGLRVSSALETTGGVSTIN | 443 |
| Db | 319 | MSMCTADQSGGIVMDSHSDNPLGAAVTLAHELGHNFNMHDTLDKGCSCQMAVEKGCIMN | 378 |
| Qy | 444 | -----PPF-----IPETPTPGMKMGKSGSGRELGE | 469 |
| Db | 379 | ASTGVPFPPMVFSSCRKDIETLSLEKGMVGCLENLEPVSESGGCKGNRVERSGE | 434 |

Search completed: April 27, 2002, 07:58:37
Job time: 738 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:55:18 ; Search time 389.72 Seconds
(without alignments)
439.840 Million cell updates/sec

Title: US-09-882-694A-5
Perfect score: 2549
Sequence: 1 MVLSPDYKSELFNNFEVFS.....EEGLKAYLEPKTINIHVTE 487

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Pending Patents AA.Main.*
- 1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep.*
 - 2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
 - 3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
 - 4: /cgn2_6/ptodata/2/paa/US080_COMB.pep.*
 - 5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
 - 6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
 - 7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
 - 8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
 - 9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
 - 10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
 - 11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
 - 12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
 - 13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
 - 14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
 - 15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
 - 16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
 - 17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
 - 18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
 - 19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
 - 20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
 - 21: /cgn2_6/ptodata/2/paa/US097_COMB.pep.*
 - 22: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
 - 23: /cgn2_6/ptodata/2/paa/US099_COMB.pep.*
 - 24: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
 - 25: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
 - 26: /cgn2_6/ptodata/2/paa/US060_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB ID | Description |
|------------|-------|-------|--------|-------|------------------|
| 1 | 2549 | 100.0 | 487 | 17 | US-09-351-224-5 |
| 2 | 2549 | 100.0 | 487 | 17 | US-09-351-224E-5 |
| 3 | 2549 | 100.0 | 487 | 17 | US-09-351-823-5 |
| 4 | 2549 | 100.0 | 487 | 20 | US-09-677-488-5 |
| 5 | 2549 | 100.0 | 487 | 20 | US-09-677-488A-5 |
| 6 | 2549 | 100.0 | 487 | 20 | US-09-677-682-5 |
| 7 | 2549 | 100.0 | 487 | 20 | US-09-677-682A-5 |

| | | | | | |
|----|-------|-------|-----|----|----------------------|
| 8 | 2549 | 100.0 | 487 | 20 | US-09-677-682B-5 |
| 9 | 2549 | 100.0 | 487 | 22 | US-09-882-694-5 |
| 10 | 2549 | 100.0 | 487 | 22 | US-09-882-694A-5 |
| 11 | 886.5 | 34.8 | 559 | 21 | US-09-760-443-1406 |
| 12 | 886.5 | 34.8 | 559 | 21 | US-09-760-475-3429 |
| 13 | 882.5 | 34.6 | 517 | 1 | PCT-US01-17253-3 |
| 14 | 882.5 | 34.6 | 517 | 19 | US-09-538-092-1047 |
| 15 | 868 | 34.1 | 497 | 10 | US-08-612-853-10 |
| 16 | 866 | 34.0 | 496 | 10 | US-08-612-853-2 |
| 17 | 866 | 34.0 | 496 | 10 | US-08-612-855-3 |
| 18 | 866 | 34.0 | 496 | 22 | US-09-847-208-59 |
| 19 | 866 | 34.0 | 496 | 22 | US-09-847-208B-59 |
| 20 | 864 | 33.9 | 497 | 10 | US-08-612-853-11 |
| 21 | 856.5 | 33.6 | 500 | 10 | US-08-612-853-15 |
| 22 | 852.5 | 33.4 | 500 | 10 | US-08-612-853-14 |
| 23 | 843.5 | 33.1 | 500 | 19 | US-09-538-092-831 |
| 24 | 840.5 | 33.0 | 517 | 19 | US-09-538-092-851 |
| 25 | 838.5 | 32.9 | 500 | 10 | US-08-612-853-16 |
| 26 | 836.5 | 32.8 | 499 | 6 | US-08-238-818-2 |
| 27 | 825.5 | 32.4 | 520 | 10 | US-08-612-853-12 |
| 28 | 824.5 | 32.3 | 512 | 19 | US-09-538-092-1139 |
| 29 | 824.5 | 32.3 | 512 | 23 | US-09-961-403-12 |
| 30 | 824.5 | 32.3 | 515 | 7 | US-08-366-210B-2 |
| 31 | 824.5 | 32.3 | 529 | 1 | PCT-US01-08631-36936 |
| 32 | 820.5 | 32.2 | 515 | 7 | US-08-366-210B-4 |
| 33 | 818.5 | 32.1 | 521 | 7 | US-08-347-326A-2 |
| 34 | 818.5 | 32.1 | 521 | 7 | US-08-347-326B-2 |
| 35 | 815.5 | 32.0 | 495 | 10 | US-08-612-855-2 |
| 36 | 815.5 | 32.0 | 495 | 22 | US-09-847-208-11 |
| 37 | 815.5 | 32.0 | 495 | 22 | US-09-847-208B-11 |
| 38 | 815.5 | 32.0 | 496 | 10 | US-08-612-853-9 |
| 39 | 808.5 | 31.7 | 500 | 10 | US-08-612-853-13 |
| 40 | 787 | 30.9 | 519 | 21 | US-09-760-443-1247 |
| 41 | 787 | 30.9 | 519 | 21 | US-09-760-475-3423 |
| 42 | 781.5 | 30.7 | 501 | 17 | US-09-344-882-22 |
| 43 | 781 | 30.6 | 538 | 17 | US-09-344-882-20 |
| 44 | 781 | 30.6 | 538 | 21 | US-09-708-427-12049 |
| 45 | 779 | 30.6 | 511 | 21 | US-09-708-427-55325 |

ALIGNMENTS

RESULT 1

US-09-351-224-5
; Sequence 5, Application US/09351224
; GENERAL INFORMATION:
; APPLICANT: Duviok, Jon
; TITLE OF INVENTION: Compositions and Methods for Fummonisin Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/351,224
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: aldehyde dehydrogenase
US-09-351-224-5

Query Match 100.0%; Score 2549; DB 17; Length 487;
Best Local Similarity 100.0%; Pred. No. 3.4e-257;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVLSPDYKSELFNNFEVSSKGSERLITNPNWDESTVATDVHVANAADVDSAVAASVQA 60
|||||
Db 1 MVLSPDYKSELFNNFEVSSKGSERLITNPNWDESTVATDVHVANAADVDSAVAASVQA 60
QY 61 VKKGPWKKTGQRRACMLKFLADLAEKNAELARLESPTGTPVSMITHEDIPNWSVER 120
|||||

Db 61 VKGPKWKTGAORAAACMLKADLAENAEKLAERLESPTGRPVSMITHFDIPNMVSVFR 120
Qy 121 YYAGWADKIAGKTFPEDNGKPNMRYEPMGVCAGIASNNATFLYVGWKIAPALAAAGCSPIF 180
Db 121 YYAGWADKIAGKTFPEDNGKPNMRYEPMGVCAGIASNNATFLYVGWKIAPALAAAGCSPIF 180
Qy 181 KASEKSPGLVGLAPLFAEAGPPGVVQFLTGTARVTGEALASHMDIAKISFTRSVGGGRA 240
Db 181 KASEKSPGLVGLAPLFAEAGPPGVVQFLTGTARVTGEALASHMDIAKISFTRSVGGGRA 240
Qy 241 VKQATLKSNNKRVTLLELGEKPTIVFNEAPLERQSGESAKDFSKFGQIWWPPSCLLVQNGN 300
Db 241 VKQATLKSNNKRVTLLELGEKPTIVFNEAPLERQSGESAKDFSKFGQIWWPPSCLLVQNGN 300
Qy 301 LAEFHGVHRGSGGCGORWLGONPLEKPTHGPFVDKSOYDRVLGNIDVGDKTQALLTGV 360
Db 301 LAEFHGVHRGSGGCGORWLGONPLEKPTHGPFVDKSOYDRVLGNIDVGDKTQALLTGV 360
Qy 361 GRKDGKGFATPTIFVNPCKPGSKIMFEEIFGVPVLSIKTFKTEEEAIEIANDTTYGLASVI 420
Db 361 GRKDGKGFATPTIFVNPCKPGSKIMFEEIFGVPVLSIKTFKTEEEAIEIANDTTYGLASVI 420
Qy 421 YTKSLNGLRVSSALETGGVSNFPIPTOTPFQGMKQSGSGRELGEGLKAYLEPKTI 480
Db 421 YTKSLNGLRVSSALETGGVSNFPIPTOTPFQGMKQSGSGRELGEGLKAYLEPKTI 480
Qy 481 NIHVNI 487
Db 481 NIHVNI 487

RESULT 2

US-09-351-224E-5

; Sequence 5, Application US/09351224E

; GENERAL INFORMATION:

; APPLICANT: Duvick, Jon

; APPLICANT: Maddox, Joyce

; APPLICANT: Gilliam, Jacob

; APPLICANT: Folkerts, Otto

; APPLICANT: Crasta, Oswald R.

; TITLE OF INVENTION: Compositions and Methods for Fumonisin

; TITLE OF INVENTION: Detoxification

; FILE REFERENCE: 5718-111

; CURRENT APPLICATION NUMBER: US/09/351,224E

; CURRENT FILING DATE: 1999-07-12

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5

; LENGTH: 487

; TYPE: PRT

; ORGANISM: Exophiala spinifera

US-09-351-224E-5

Query Match 100.0%; Score 2549; DB 17; Length 487;
Best Local Similarity 100.0%; Pred. No. 3.4e-257;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVLSPEYKSELFINNEFVSSKGSERLTLTNPDWDESTVATDVHVAADVDVAASVQA 60
Db 1 MVLSPEYKSELFINNEFVSSKGSERLTLTNPDWDESTVATDVHVAADVDVAASVQA 60
Qy 61 VKGPKWKTGAORAAACMLKADLAENAEKLAERLESPTGRPVSMITHFDIPNMVSVFR 120
Db 61 VKGPKWKTGAORAAACMLKADLAENAEKLAERLESPTGRPVSMITHFDIPNMVSVFR 120
Qy 121 YYAGWADKIAGKTFPEDNGKPNMRYEPMGVCAGIASNNATFLYVGWKIAPALAAAGCSPIF 180
Db 121 YYAGWADKIAGKTFPEDNGKPNMRYEPMGVCAGIASNNATFLYVGWKIAPALAAAGCSPIF 180
Qy 181 KASEKSPGLVGLAPLFAEAGPPGVVQFLTGTARVTGEALASHMDIAKISFTRSVGGGRA 240
Db 181 KASEKSPGLVGLAPLFAEAGPPGVVQFLTGTARVTGEALASHMDIAKISFTRSVGGGRA 240

QY 421 YTKSLNRLRVSSALETGCGVSINFFPIPETOTPPGGMKQSGSGRELGEGLKAYLEPKTI 480
Db 421 YTKSLNRLRVSSALETGCGVSINFFPIPETOTPPGGMKQSGSGRELGEGLKAYLEPKTI 480
QY 481 NIHVNIE 487
Db 481 NIHVNIE 487

RESULT 4
US-09-677-488-5
; Sequence 5, Application US/09677488
; GENERAL INFORMATION:
; APPLICANT: Duwrick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin Detoxification
; FILE REFERENCE: 35718/204100
; CURRENT APPLICATION NUMBER: US/09/677,488A
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: aldehyde dehydrogenase, translation of fully
; OTHER INFORMATION: spliced cdna
US-09-677-488-5

Query Match 100.0%; Score 2549; DB 20; Length 487;
Best Local Similarity 100.0%; Pred. No. 3.4e-257;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVLSPDEYKSELFINNEFVSSKGSRLTLTNPWDESTVATDVHANAADVDSAAVSAQA 60
Db 1 MVLSPDEYKSELFINNEFVSSKGSRLTLTNPWDESTVATDVHANAADVDSAAVSAQA 60
QY 61 VKKGPWKFTGAQRAACMLKPADLAENAKLAEKLALESPTGRPVSMITHDIPNMVSVFR 120
Db 61 VKKGPWKFTGAQRAACMLKPADLAENAKLAEKLALESPTGRPVSMITHDIPNMVSVFR 120
QY 121 YAGWADKIAGTFFEDNGKPNWRYEPMGVCAAGIASNNATFLYGVWKIAPALAAAGCSFIF 180
Db 121 YAGWADKIAGTFFEDNGKPNWRYEPMGVCAAGIASNNATFLYGVWKIAPALAAAGCSFIF 180
QY 181 KASEKSPGLVGLAPLFAEAGFPVGVQVFLTCARVTGEALASHMDIAKISFTSVSGGGA 240
Db 181 KASEKSPGLVGLAPLFAEAGFPVGVQVFLTCARVTGEALASHMDIAKISFTSVSGGGA 240
QY 241 VKOATLKSNNKRVTLLEGEKPTIVFNEAPLERQSGESAKDFSKFGQIWVPPSCLLVQWGN 300
Db 241 VKOATLKSNNKRVTLLEGEKPTIVFNEAPLERQSGESAKDFSKFGQIWVPPSCLLVQWGN 300
QY 301 LAEKFHGVRHSGFCGQWLGONPLEKPRTHGPFVDKSOYDRVLGNIDVGKDTAQLLTGV 360
Db 301 LAEKFHGVRHSGFCGQWLGONPLEKPRTHGPFVDKSOYDRVLGNIDVGKDTAQLLTGV 360
QY 361 GRKGDGFAIEPTIFVNPKPGSKIWFEIFGPVLSIKFTKTEEEAIEIANDTTYGLASVI 420
Db 361 GRKGDGFAIEPTIFVNPKPGSKIWFEIFGPVLSIKFTKTEEEAIEIANDTTYGLASVI 420
QY 421 YTKSLNRLRVSSALETGCGVSINFFPIPETOTPPGGMKQSGSGRELGEGLKAYLEPKTI 480
Db 421 YTKSLNRLRVSSALETGCGVSINFFPIPETOTPPGGMKQSGSGRELGEGLKAYLEPKTI 480
QY 481 NIHVNIE 487
Db 481 NIHVNIE 487

RESULT 5
US-09-677-488A-5
; Sequence 5, Application US/09677488A
; GENERAL INFORMATION:
; APPLICANT: Duwrick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 35718/204100
; CURRENT APPLICATION NUMBER: US/09/677,488A
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-677-488A-5

Query Match 100.0%; Score 2549; DB 20; Length 487;
Best Local Similarity 100.0%; Pred. No. 3.4e-257;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVLSPDEYKSELFINNEFVSSKGSRLTLTNPWDESTVATDVHANAADVDSAAVSAQA 60
Db 1 MVLSPDEYKSELFINNEFVSSKGSRLTLTNPWDESTVATDVHANAADVDSAAVSAQA 60
QY 61 VKKGPWKFTGAQRAACMLKPADLAENAKLAEKLALESPTGRPVSMITHDIPNMVSVFR 120
Db 61 VKKGPWKFTGAQRAACMLKPADLAENAKLAEKLALESPTGRPVSMITHDIPNMVSVFR 120
QY 121 YAGWADKIAGTFFEDNGKPNWRYEPMGVCAAGIASNNATFLYGVWKIAPALAAAGCSFIF 180
Db 121 YAGWADKIAGTFFEDNGKPNWRYEPMGVCAAGIASNNATFLYGVWKIAPALAAAGCSFIF 180
QY 181 KASEKSPGLVGLAPLFAEAGFPVGVQVFLTCARVTGEALASHMDIAKISFTSVSGGGA 240
Db 181 KASEKSPGLVGLAPLFAEAGFPVGVQVFLTCARVTGEALASHMDIAKISFTSVSGGGA 240
QY 241 VKOATLKSNNKRVTLLEGEKPTIVFNEAPLERQSGESAKDFSKFGQIWVPPSCLLVQWGN 300
Db 241 VKOATLKSNNKRVTLLEGEKPTIVFNEAPLERQSGESAKDFSKFGQIWVPPSCLLVQWGN 300
QY 301 LAEKFHGVRHSGFCGQWLGONPLEKPRTHGPFVDKSOYDRVLGNIDVGKDTAQLLTGV 360
Db 301 LAEKFHGVRHSGFCGQWLGONPLEKPRTHGPFVDKSOYDRVLGNIDVGKDTAQLLTGV 360
QY 361 GRKGDGFAIEPTIFVNPKPGSKIWFEIFGPVLSIKFTKTEEEAIEIANDTTYGLASVI 420
Db 361 GRKGDGFAIEPTIFVNPKPGSKIWFEIFGPVLSIKFTKTEEEAIEIANDTTYGLASVI 420
QY 421 YTKSLNRLRVSSALETGCGVSINFFPIPETOTPPGGMKQSGSGRELGEGLKAYLEPKTI 480
Db 421 YTKSLNRLRVSSALETGCGVSINFFPIPETOTPPGGMKQSGSGRELGEGLKAYLEPKTI 480
QY 481 NIHVNIE 487
Db 481 NIHVNIE 487

RESULT 6
US-09-677-682-5
; Sequence 5, Application US/09677682
; GENERAL INFORMATION:
; APPLICANT: Duwrick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fumonisin Detoxification

```
; FILE REFERENCE: 35718/204101 (5718-111B )
; CURRENT APPLICATION NUMBER: US/09/677,682
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: aldehyde dehydrogenase, translation of fully
; OTHER INFORMATION: spliced cDNA
US-09-677-682-5

Query Match      100.0%; Score 2549; DB 20; Length 487;
Best Local Similarity 100.0%; Pred. No. 3.4e-257;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVLSPDEYKSELFINNEFVSSKSGSERLTLTNPWDESTVATDVHVAADVDVAASVQA 60
    |||||||
Db 1 MVLSPDEYKSELFINNEFVSSKSGSERLTLTNPWDESTVATDVHVAADVDVAASVQA 60
Qy 61 VKKGPWKFTGAQRAACMLKFAADLAENAEKARLESPTGRPVSMITHFDIPNMVSVFR 120
    |||||||
Db 61 VKKGPWKFTGAQRAACMLKFAADLAENAEKARLESPTGRPVSMITHFDIPNMVSVFR 120
Qy 121 YYAGWADKIAGKTFPEDNGKPNWRYEPMGVCAGIASNNATFLYVGWKIAPALAAAGCSFIF 180
    |||||||
Db 121 YYAGWADKIAGKTFPEDNGKPNWRYEPMGVCAGIASNNATFLYVGWKIAPALAAAGCSFIF 180
Qy 181 KASEKSPGLVGLAPLFAEAGFPVGVQFLTGARVTGEALASHMDIAKISFTRSVGGGRA 240
    |||||||
Db 181 KASEKSPGLVGLAPLFAEAGFPVGVQFLTGARVTGEALASHMDIAKISFTRSVGGGRA 240
Qy 241 VKQATLSNMRKRYTLELGEKPTIVFNEAPLEROSGESAKDFSKFGQIWWPPSCLLVQWGN 300
    |||||||
Db 241 VKQATLSNMRKRYTLELGEKPTIVFNEAPLEROSGESAKDFSKFGQIWWPPSCLLVQWGN 300
Qy 301 LAEFHGVHSGSGCGQRLGQNPLEPKRTHGPFVDKSOYDRVLGNIDVGKDTAQLLTGV 360
    |||||||
Db 301 LAEFHGVHSGSGCGQRLGQNPLEPKRTHGPFVDKSOYDRVLGNIDVGKDTAQLLTGV 360
Qy 361 GRKDGKGFALPTIFVNPKPGSKINFEELFGPVLSTIKTFKEEAEIANDTTYGLASVI 420
    |||||||
Db 361 GRKDGKGFALPTIFVNPKPGSKINFEELFGPVLSTIKTFKEEAEIANDTTYGLASVI 420
Qy 421 YTKSLNRLRVSSALETGGVSINFPFIPETQTPFGMKQSGSGRELGEGLKAYLEPKTI 480
    |||||||
Db 421 YTKSLNRLRVSSALETGGVSINFPFIPETQTPFGMKQSGSGRELGEGLKAYLEPKTI 480
Qy 481 NIHWNIE 487
    |||||||
Db 481 NIHWNIE 487

RESULT 7
US-09-677-682a-5
; Sequence 5, Application US/09677682a
; GENERAL INFORMATION:
; APPLICANT: Duivick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 35718/204101
; CURRENT APPLICATION NUMBER: US/09/677,682B
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Exophiala spinifera
```

```
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-677-682A-5

Query Match      100.0%; Score 2549; DB 20; Length 487;
Best Local Similarity 100.0%; Pred. No. 3.4e-257;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVLSPDEYKSELFINNEFVSSKSGSERLTLTNPWDESTVATDVHVAADVDVAASVQA 60
    |||||||
Db 1 MVLSPDEYKSELFINNEFVSSKSGSERLTLTNPWDESTVATDVHVAADVDVAASVQA 60
Qy 61 VKKGPWKFTGAQRAACMLKFAADLAENAEKARLESPTGRPVSMITHFDIPNMVSVFR 120
    |||||||
Db 61 VKKGPWKFTGAQRAACMLKFAADLAENAEKARLESPTGRPVSMITHFDIPNMVSVFR 120
Qy 121 YYAGWADKIAGKTFPEDNGKPNWRYEPMGVCAGIASNNATFLYVGWKIAPALAAAGCSFIF 180
    |||||||
Db 121 YYAGWADKIAGKTFPEDNGKPNWRYEPMGVCAGIASNNATFLYVGWKIAPALAAAGCSFIF 180
Qy 181 KASEKSPGLVGLAPLFAEAGFPVGVQFLTGARVTGEALASHMDIAKISFTRSVGGGRA 240
    |||||||
Db 181 KASEKSPGLVGLAPLFAEAGFPVGVQFLTGARVTGEALASHMDIAKISFTRSVGGGRA 240
Qy 241 VKQATLSNMRKRYTLELGEKPTIVFNEAPLEROSGESAKDFSKFGQIWWPPSCLLVQWGN 300
    |||||||
Db 241 VKQATLSNMRKRYTLELGEKPTIVFNEAPLEROSGESAKDFSKFGQIWWPPSCLLVQWGN 300
Qy 301 LAEFHGVHSGSGCGQRLGQNPLEPKRTHGPFVDKSOYDRVLGNIDVGKDTAQLLTGV 360
    |||||||
Db 301 LAEFHGVHSGSGCGQRLGQNPLEPKRTHGPFVDKSOYDRVLGNIDVGKDTAQLLTGV 360
Qy 361 GRKDGKGFALPTIFVNPKPGSKINFEELFGPVLSTIKTFKEEAEIANDTTYGLASVI 420
    |||||||
Db 361 GRKDGKGFALPTIFVNPKPGSKINFEELFGPVLSTIKTFKEEAEIANDTTYGLASVI 420
Qy 421 YTKSLNRLRVSSALETGGVSINFPFIPETQTPFGMKQSGSGRELGEGLKAYLEPKTI 480
    |||||||
Db 421 YTKSLNRLRVSSALETGGVSINFPFIPETQTPFGMKQSGSGRELGEGLKAYLEPKTI 480
Qy 481 NIHWNIE 487
    |||||||
Db 481 NIHWNIE 487

RESULT 8
US-09-677-682B-5
; Sequence 5, Application US/09677682B
; GENERAL INFORMATION:
; APPLICANT: Duivick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 35718/204101
; CURRENT APPLICATION NUMBER: US/09/677,682B
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Exophiala spinifera
```

US-09-677-682B-5

Query Match 100.0%; Score 2549; DB 20; Length 487;
Best Local Similarity 100.0%; Pred. No. 3.4e-257;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVLSPDEYKSELFNNFVSSKGSRLTLTNPDWDESTVATDVHANAADVDVAASVQA 60
DB 1 MVLSPDEYKSELFNNFVSSKGSRLTLTNPDWDESTVATDVHANAADVDVAASVQA 60
QY 61 VKKGPWKFTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSVFR 120
DB 61 VKKGPWKFTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSVFR 120
QY 121 YYAGWADKTAGKTFPEDNGKPNWRYPGVCAGIASWNATFLYVGWKIAPALAAACCSFIF 180
DB 121 YYAGWADKTAGKTFPEDNGKPNWRYPGVCAGIASWNATFLYVGWKIAPALAAACCSFIF 180
QY 181 KASEKSPGLVGLIAPLFAEAGPPGVVQFLTGARVTGEALASHMDIAKISFTRSVGGGRA 240
DB 181 KASEKSPGLVGLIAPLFAEAGPPGVVQFLTGARVTGEALASHMDIAKISFTRSVGGGRA 240
QY 241 VKOATLKSNNKRVTLLEGEKPTIVNEAPLEROSGESAKDFSKFGQIWVPPSCLLVQWGN 300
DB 241 VKOATLKSNNKRVTLLEGEKPTIVNEAPLEROSGESAKDFSKFGQIWVPPSCLLVQWGN 300
QY 301 LAEKFGVHRHSGFGGCGQRLGNPLPKRTHGPFVDKSOYDRVLGNIDVGDKTAQLLTGV 360
DB 301 LAEKFGVHRHSGFGGCGQRLGNPLPKRTHGPFVDKSOYDRVLGNIDVGDKTAQLLTGV 360
QY 361 GRGDKGFALPTIFVNPKPGSKIMFEEIFGPVLSIKTFKTEEEAIEIANDTTYGLASVI 420
DB 361 GRGDKGFALPTIFVNPKPGSKIMFEEIFGPVLSIKTFKTEEEAIEIANDTTYGLASVI 420
QY 421 YTKSLNRLRVSSALETTGGVSNFPIPETQTPFGGMKQSGRGELGEGLKAYLEPKTI 480
DB 421 YTKSLNRLRVSSALETTGGVSNFPIPETQTPFGGMKQSGRGELGEGLKAYLEPKTI 480
QY 481 NIHVNIE 487
DB 481 NIHVNIE 487

RESULT 9

US-09-882-694-5

; Sequence 5, Application US/09882694
; GENERAL INFORMATION:
; APPLICANT: Duwick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 35718/208255
; CURRENT APPLICATION NUMBER: US/09/882,694
; PRIOR FILING DATE: 2001-06-12
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-882-694-5

Query Match 100.0%; Score 2549; DB 22; Length 487;
Best Local Similarity 100.0%; Pred. No. 3.4e-257;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVLSPDEYKSELFNNFVSSKGSRLTLTNPDWDESTVATDVHANAADVDVAASVQA 60
DB 1 MVLSPDEYKSELFNNFVSSKGSRLTLTNPDWDESTVATDVHANAADVDVAASVQA 60
QY 61 VKKGPWKFTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSVFR 120
DB 61 VKKGPWKFTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSVFR 120
QY 121 YYAGWADKTAGKTFPEDNGKPNWRYPGVCAGIASWNATFLYVGWKIAPALAAACCSFIF 180
DB 121 YYAGWADKTAGKTFPEDNGKPNWRYPGVCAGIASWNATFLYVGWKIAPALAAACCSFIF 180
QY 181 KASEKSPGLVGLIAPLFAEAGPPGVVQFLTGARVTGEALASHMDIAKISFTRSVGGGRA 240
DB 181 KASEKSPGLVGLIAPLFAEAGPPGVVQFLTGARVTGEALASHMDIAKISFTRSVGGGRA 240
QY 241 VKOATLKSNNKRVTLLEGEKPTIVNEAPLEROSGESAKDFSKFGQIWVPPSCLLVQWGN 300
DB 241 VKOATLKSNNKRVTLLEGEKPTIVNEAPLEROSGESAKDFSKFGQIWVPPSCLLVQWGN 300
QY 301 LAEKFGVHRHSGFGGCGQRLGNPLPKRTHGPFVDKSOYDRVLGNIDVGDKTAQLLTGV 360
DB 301 LAEKFGVHRHSGFGGCGQRLGNPLPKRTHGPFVDKSOYDRVLGNIDVGDKTAQLLTGV 360
QY 361 GRGDKGFALPTIFVNPKPGSKIMFEEIFGPVLSIKTFKTEEEAIEIANDTTYGLASVI 420
DB 361 GRGDKGFALPTIFVNPKPGSKIMFEEIFGPVLSIKTFKTEEEAIEIANDTTYGLASVI 420
QY 421 YTKSLNRLRVSSALETTGGVSNFPIPETQTPFGGMKQSGRGELGEGLKAYLEPKTI 480
DB 421 YTKSLNRLRVSSALETTGGVSNFPIPETQTPFGGMKQSGRGELGEGLKAYLEPKTI 480
QY 481 NIHVNIE 487
DB 481 NIHVNIE 487

RESULT 10

US-09-882-694A-5

; Sequence 5, Application US/09882694A
; GENERAL INFORMATION:
; APPLICANT: Duwick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 3718-111
; CURRENT APPLICATION NUMBER: US/09/882,694A
; PRIOR FILING DATE: 2001-06-15
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-882-694A-5

Query Match 100.0%; Score 2549; DB 22; Length 487;
Best Local Similarity 100.0%; Pred. No. 3.4e-257;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVLSPDEYKSELFNNFVSSKGSRLTLTNPDWDESTVATDVHANAADVDVAASVQA 60
DB 1 MVLSPDEYKSELFNNFVSSKGSRLTLTNPDWDESTVATDVHANAADVDVAASVQA 60
QY 61 VKKGPWKFTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSVFR 120
DB 61 VKKGPWKFTGAQRAACMLKFADLAEKNAEKLARLESPTGRPVSMITHFDIPNMVSVFR 120

Db 372 GYIOLGQKACALLCGGERGFFIKPTVFGVQDDMRIAKEEIFGPPVQPLFKFKKIE 431
QY 404 EATEIANDTTYGLASVIYTKSLNRLVSSALETGGVSINFPFIPETOTPFPGCKMGSGG 463
Db 432 EVVERANTRYGLAAAVFTDLDKAMYFTQALQAGTVWNTYINIVTCHTTPFGGFKESGNG 491
QY 464 RELGEGLKAYLEPKTINIHV 484
Db 492 RELGEDGLKAYTEVKVITIKV 512

RESULT 15
US-08-612-853-10
; Sequence 10, Application US/08612853
; GENERAL INFORMATION:
; APPLICANT: Achatz, Gernot
; APPLICANT: Oberkofler, Hannes
; APPLICANT: Simon, Birgit
; APPLICANT: Unger, Andrea
; APPLICANT: Lechenauer, Erich
; APPLICANT: Hirschwehr, Reinhold
; TITLE OF INVENTION: Recombinant Cladosporium Herbarum
; TITLE OF INVENTION: Allergens
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,853
; FILING DATE: 26-FEB-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AT94/00120
; FILING DATE: 02-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Friebe, Thomas E
; REGISTRATION NUMBER: 29,258
; REFERENCE/DOCKET NUMBER: 6530-021-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 497 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Cladosporium herbarum
US-08-612-853-10

Query Match 34.1%; Score 868; DB 10; Length 497;
Best Local Similarity 42.8%; Pred No. 5.3e-81;
Matches 211; Conservative 67; Mismatches 183; Indels 32; Gaps 12;
QY 7 EYKSELFINNEFVSSKGSERLTLTNPWDESTVATDVHVAADVDVSAVAASVQAVKGPW 66
Db 15 EQPTGLFINNEFVKGEGKTFDVPINPDES-VITQVHEATEKDVDIAVAARQAF-EGSW 72
QY 67 KKFTGAQRAACMLKPADLAEKNAEKLALESPTGRPVSMITHFDIPNMVSV-----FRY 121

Db 73 RLETPENRGKLLNNLANLFENKNTDLLAAVESLONGKATSM-----ARVTSACASGCLRY 136
QY 122 YAGWADKIAGKTFPEBNGRPNHRY----EPMGVCAGIASWNATFLYVGVKWIAPALAACCSF 178
Db 127 YGWMADKIITGKVI--DTTPTDNTNYVKKEPIGVCRSDHLSLEPLPLMMWAMKIGPAIAGNTV 184
QY 179 IFKASEKSPLGVLGLAPFAEAGFPVGVQFLTG-ARVTGEALASHMDIAKISFTKSVGG 237
Db 185 VLKTAEQTPGLGLVAASLVKEAGFPVGVINVISGFGKVAGAAALSSHMDVDKVAFTGTVV 244
QY 238 GRAVKQATLKSNKRVTLLEGEK-PTIVNEAPLERQ-SGESAKDFSKFGQIWPVPSCLL 295
Db 245 GRTILKAAAASSNLKVTLELGCKSPNIVFEDADIDNAISWVNFGIFNFHGGCCCCAGSRVY 304
QY 296 VQ---WGNLAERFHGVHRHSGFGGCRWLGNQNPLEPKRTHGPFVDRKQYDRLGNIDVQKD 352
Db 305 VQESIYDKFVQKFK-----EKAQNVVDGDPFAADTFQGPQVSKYQFDRIMEYIQAGKD 357
QY 353 T-AQLLTGVRKGDGKGFALIEPTIFVNPKNPKGSKIWFEEIFGPVLSIKTKFTKEBAIAND 411
Db 358 AGATVETGSGRKGDKGYFTIEPTIFSNVTEDMKIVKEEIEFGVPCVSIKAKFKTKEDAIKLGNA 417
QY 412 TTYGLASVIYTKSLNRLVSSALETGGVSINFPFIPETOTPFPGMKQSGSGRELGEGL 471
Db 418 STYGLAAAVHTKNLNTAIEVSNALKAGTYWNTYHTLHHQMPFGGYKESGIGRELGEDAL 477
QY 472 KAYLEPKTINIHV 484
Db 478 ANYTQTKTVSIRL 490

Search completed: April 27, 2002, 07:55:20
Job time: 581 sec

```
QY 69 FTGQRAACMLKPADLAENAEKARLESPT---GRPVSMITHFDIPNNWSVFRYYAGW 125
Db 96 MDASERGLYKKADLIIRD-RLLATMESMGNGKLYSNAYLNDLAGCIKTLRYCAGW 154
QY 126 ADKI--AGKTFPED-NGKPNWRYEPMGVGAGIASNNATPLYVGWKIAPALAAAGCSFIFKA 182
Db 155 ADKIOGQRTIPIDGNFTYTRHEPIGVCGQIIPKNFPLVMLIKIPALSGNIVVYKP 214
QY 183 SEKSPGLVGLAPLAFAGPPGVVOLTGARVT-GEALASHMDIAKISFTRSVGGGRAY 241
Db 215 AEQPLTALHVASLIKAGPPGVNIVPGYGTAGAAISSHMDIKVAFGTGSTVGLKI 274
QY 242 KOATLKNMKRVLTLEGEK-PTTVFNAPLERQSGESAKDESKFGQIWWPPSCLLVQNGN 300
Db 275 KEAAGKSNLKRVTLELGGKSPCLVADADL-----DNAVEFAHGVVYHQGCCIAASRI 329
QY 301 LAEK--PHGVHSGFCGCCRWLQONLEPKRTHGPPVDSQDRVLGNIDVG-KDTAQLL 357
Db 330 FVEESIYDEFVRSVERAKYIILGNLTPGVTOGPOIDKEQYDKILOLLESKKEGAKLE 389
QY 358 TGVGRKDGKFAIRPTIFVNPKPGSKIWFEEIFPGPVLSIKTFKTEPEAEIANDTYGLA 417
Db 390 CGGPGWNGKYFVQPTVFSNVTDEMRIAKEEIFPGPVQOIMKFKSLDDVTKRANNTFYGLS 449
QY 418 SVIYTKSLNGLRVSSALETGGVSIINFPFIPETQTPFGGMKOSGSGRELGEGLKAYLEP 477
Db 450 AGVETKIDRAITISSALQAGTVWNCYGVVSAQCFPGGFKMSGNGRELGEVGFHEYTEV 509
QY 478 KTIINHVN 485
Db 510 KTVTVKIS 517
```

RESULT 2

```
US-08-513-841-2
; Sequence 2, Application US/08513841
; Patent No. 5753481
; GENERAL INFORMATION:
; APPLICANT: Niwa, Mineo
; APPLICANT: Saito, Yoshimasa
; APPLICANT: Ishii, Yoshinori
; APPLICANT: Yoshida, Masaru
; APPLICANT: Suzuki, Hiromi
; TITLE OF INVENTION: No. 5753481el L-sorbos Dehydrogenase and No. 5753481el L-sorbo
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESS: Obion, Spivak, McClelland, Maier & Neustadt, P.C.
; STREET: 1755 Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS-DOS Editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/513,841
; FILING DATE: 01-NOV-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 9304700.9
; FILING DATE: 08-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 241851/1993
; FILING DATE: 28-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-909-0 PCT
; TELECOMMUNICATION INFORMATION:
```

```
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 497 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Gluconobacter oxydans
; STRAIN: T-100
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: 1..497
; IDENTIFICATION METHOD: experimentally
; US-08-513-841-2

Query Match . 24.0%; Score 612.5; DB 1; Length 497;
Best Local Similarity 34.6%; Pred. No. 4.4e-60;
Matches 174; Conservative 83; Mismatches 197; Indels 49; Gaps 14;

QY 3 LSPDEYKSELFTNNEPVSSKSERLTLTNPWDESVATDVHV-----ANAAADVSAVAAS 57
Db 11 LKPREP--GFFIDGHWRAK-----DFFDRSSPAHDVPVTRIPRCTREDLDEAVAAA 60
QY 58 VQAVKKGPMKKFTGQRAACMLKPADLAENAEKARLESPTGRPVSMITHFDIPNNWS 117
Db 61 RRAFENGWAGLAADRAAVALLLKAGLLRRERDDIAYWEVLKNGKPISQ-AKGEIDHCIA 119
QY 118 VFRYVAGWADKIAGTTFPEDNGKPNWRY-----EPMGVCAAGIASNNATFLYVGWKIAPAL 172
Db 120 CFEMAAGAARMLRGDTF-----NNLGEGLGWLREPVGVLITPWNPPFMCILCERAPPIL 176
QY 173 AACCSFIFKASEKSPGLVGLAPLAFAGPPGVVQVLTG-ARTVGEALASHMDIAKISF 231
Db 177 ASGCTLVVRPAEVTSAITLLAEIADADAGLPKGVNVTGTRVVGQANTHEQDIDMLSF 236
QY 232 TRSVGGGRVAKQATLKNMKRVLTLEGEK-PTTVFNAPLERQSGESAKDFS-KFGQIWW 289
Db 237 TGSTGVGKSCIHAADSNLKLGLGELGKNPIVVPADSNLEDAADAVAFGFSFNTGQCCV 296
QY 290 PPSCLLVQNGNLAEKF-----HGVHRHSGSGCGCORWLGONPLEPKRTHGPFVDKSOYD 341
Db 297 SSSRLIVE-RSVAEKFERLVVPKMEKIRVG-----DPPDPETQIGAITTEAQNK 344
QY 342 RVLGNTDVOK-DTAQLLTGVGRKG-DKGFAIEPTIFVNPKPGSKIWFEEIFGPVLSIKTF 399
Db 345 TILDYIAKGAEGAKLLCGGIVDFGKGYIQPTLFTDVKPMSGIARDEIFGPVLASFHF 404
QY 400 KTEEAIEIANDTYGLASVIYTKSLNGLRVSSALETGGVSIINFPFIPETQTPFGGMKQ 459
Db 405 DTVDEAIAIANDTVGLAASVWSKDIKALAVTRVRAGREWWNTIMSGGPETPLGGFKQ 464
QY 460 SSGSRELGEGLKAYLEPKTINI 482
Db 465 SCWGREGALYGVVEYTOIKSVHI 487

RESULT 3
; US-08-696-834-2
; Sequence 2, Application US/08696834
; Patent No. 5834283
; GENERAL INFORMATION:
; APPLICANT: Niwa, Mineo
; APPLICANT: Saito, Yoshimasa
; APPLICANT: Ishii, Yoshinori
; APPLICANT: Yoshida, Masaru
; APPLICANT: Hayashi, Hiromi
; TITLE OF INVENTION: Method for Producing 2-Keto-L-Gulonic Acid
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
```


Query Match 24.0%; Score 612.5; DB 2; Length 497;
Best Local Similarity 34.6%; Pred. No. 4.4e-60;
Matches 174; Conservative 83; Mismatches 197; Indels 49; Gaps 14;

QY 3 LSPDEYKSELEFINNEFVSSKSESLITLNPWDESTVATDVHV-----ANADVDVSAVAAS 57
DB 11 LKPREF--GFFIDENRAGK-----DFFRSSPAHDVPTVTRPCTREDLDVAANA 60

QY 58 VOAYKGPWKFTGAQACMLPADLAENAKELARLESPTGRPVSMITHFDIPNNVS 117
DB 61 RRAFENGWAGLAADRAAALLKAGLLRERRDDIAYWEVLENGKPI50-ARKEIDHICIA 119

QY 118 VFRYAGWADKIAGKTFFEDNGKPNRY-----EPMGVGAGIASWNAFLYVGMKIAPAL 172
DB 120 CFEMAGAAARMHGDTE---NNLCEGLFGMVLREPIGVVGLITPNWPFMILCERAPPIL 176

QY 173 AAGCSFIFKASEKSPGLVGLAPLAFAGPPGVVQFLTG-ARVTGEALASHMDIAKTSF 231
DB 177 ASGCTLVVYKPAEYTSATLLAEILADAGLPKGVFNVTGTGTGTQVQAMTEHODIMLSF 236

QY 232 TRSVGGGAVKQATLKSNMKTVELEGEK-PTIVFNEAPLERQSGESAKDFS-KFGQIWW 289
DB 237 TGSTGVGKSCIIHAADSNAKKGLEGGKNPVIVFADSNLEDAADAVAFISFNTGQCCV 296

QY 290 PFCILLVQWGNLAEKF-----HGVHSGFCGQCORWLGONPLEPKRTHGPFVDKSOYD 341
DB 297 SSSRLIVE-RSVAEKFERLVVPMKERTVIG-----DFFDPETOIGAITEAQNK 344

QY 342 RVLGNIDVGK-DTAQLLTGVGRKG-DKGFAIEPTIFVNPKNPKSKIWFEEIFGPVLSIKTF 399
DB 345 TILDYIAGKAEGAKLLCGGIVDFGKGGOYIQTFTDVKPMSGIARDEIFGPVLSAHPF 404

QY 400 KTEEAETIANDTYGLASVLYTKSLNRLVSSALETGVSINFPFIPETOTPPGGMKO 459
DB 405 DTVDIAIAIANDTYGLAASVWSKDIKALAVTRVRAGRFWNTIMSGGPTPLGGFKQ 464

QY 460 SGSGRELGEGLKAYLEPKTINI 482
DB 465 SCWGREGAGLYVEEYTIQKSVHI 487

RESULT 5
US-09-118-317-2
Sequence 2, Application US/09118317
Patent No. 6197562
GENERAL INFORMATION:
APPLICANT: Niwa, Mineo
APPLICANT: Saito, Yoshinasa
APPLICANT: Ishii, Yoshinori
APPLICANT: Yoshida, Masaru
APPLICANT: Suzuki, Hiromi
TITLE OF INVENTION: No. 6197562el L-sorbose Dehydrogenase and No. 6197562el
TITLE OF INVENTION: L-sorbose Dehydrogenase Obtained from Gluconobacter
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESSES:
ADDRESS: Obion, Spivak, McClelland, Maier & Neustadt, P.C.
STREET: 1755 Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MS-DOS Editor
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/118,317
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/513,841

FILING DATE: 01-NOV-1995
APPLICATION NUMBER: UK 9304700.9
FILING DATE: 08-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 241851/1993
FILING DATE: 28-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: NORMAN F. OBLON
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 18-909-0 PCT
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Gluconobacter oxydans
STRAIN: T-100
FEATURE:
NAME/KEY: mat peptide
LOCATION: 1..497
IDENTIFICATION METHOD: experimentally
US-09-118-317-2

Query Match 24.0%; Score 612.5; DB 4; Length 497;
Best Local Similarity 34.6%; Pred. No. 4.4e-60;
Matches 174; Conservative 83; Mismatches 197; Indels 49; Gaps 14;

QY 3 LSPDEYKSELEFINNEFVSSKSESLITLNPWDESTVATDVHV-----ANADVDVSAVAAS 57
DB 11 LKPREF--GFFIDENRAGK-----DFFRSSPAHDVPTVTRPCTREDLDVAANA 60

QY 58 VOAYKGPWKFTGAQACMLPADLAENAKELARLESPTGRPVSMITHFDIPNNVS 117
DB 61 RRAFENGWAGLAADRAAALLKAGLLRERRDDIAYWEVLENGKPI50-ARKEIDHICIA 119

QY 118 VFRYAGWADKIAGKTFFEDNGKPNRY-----EPMGVGAGIASWNAFLYVGMKIAPAL 172
DB 120 CFEMAGAAARMHGDTE---NNLCEGLFGMVLREPIGVVGLITPNWPFMILCERAPPIL 176

QY 173 AAGCSFIFKASEKSPGLVGLAPLAFAGPPGVVQFLTG-ARVTGEALASHMDIAKTSF 231
DB 177 ASGCTLVVYKPAEYTSATLLAEILADAGLPKGVFNVTGTGTGTQVQAMTEHODIMLSF 236

QY 232 TRSVGGGAVKQATLKSNMKTVELEGEK-PTIVFNEAPLERQSGESAKDFS-KFGQIWW 289
DB 237 TGSTGVGKSCIIHAADSNAKKGLEGGKNPVIVFADSNLEDAADAVAFISFNTGQCCV 296

QY 290 PFCILLVQWGNLAEKF-----HGVHSGFCGQCORWLGONPLEPKRTHGPFVDKSOYD 341
DB 297 SSSRLIVE-RSVAEKFERLVVPMKERTVIG-----DFFDPETOIGAITEAQNK 344

QY 342 RVLGNIDVGK-DTAQLLTGVGRKG-DKGFAIEPTIFVNPKNPKSKIWFEEIFGPVLSIKTF 399
DB 345 TILDYIAGKAEGAKLLCGGIVDFGKGGOYIQTFTDVKPMSGIARDEIFGPVLSAHPF 404

QY 400 KTEEAETIANDTYGLASVLYTKSLNRLVSSALETGVSINFPFIPETOTPPGGMKO 459
DB 405 DTVDIAIAIANDTYGLAASVWSKDIKALAVTRVRAGRFWNTIMSGGPTPLGGFKQ 464

QY 460 SGSGRELGEGLKAYLEPKTINI 482
DB 465 SCWGREGAGLYVEEYTIQKSVHI 487

RESULT 6
US-09-651-941-9

```
; Sequence 9, Application US/09651941
; Patent No. 6355470
; GENERAL INFORMATION:
; APPLICANT: ROUVIER, PIERRE E
; APPLICANT: WALTERS, DANA M
; APPLICANT: RAINIER, RUSS
; TITLE OF INVENTION: Genes Encoding Plicic Acid Degradation
; FILE REFERENCE: BC1022 US NA
; CURRENT APPLICATION NUMBER: US/09/651,941
; CURRENT FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,545
; PRIOR FILING DATE: 1999-10-03
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Rhodococcus erythropolis HL PM-1
; US-09-651-941-9

Query Match      23.8%; Score 605.5; DB 4; Length 485;
Best Local Similarity 32.3%; Pred. No. 2.6e-59;
Matches 159; Conservative 81; Mismatches 206; Indels 47; Gaps 11;

QY 12 LFINFEVSSKGSERLTLTNPMDESTVATDVHVAADVDSDAASVAQVKKGPWKFTG 71
DB 10 LVIGDQLTPSSTGATFDSINPADGSHLAS-VAEATAADVARAVEAAKAAAR--TWQMRP 66
QY 72 AORACMLKPADLAEKNAEKLESPTGRPVSMITHFDIPNMVSVFRYYAGNADKIAG 131
DB 67 AORTLMFRYAALIEBEKTELAQSDMDGKPIRESLGIDLPIIMETLEYFAGLVTKIEG 126
QY 132 KTFPEDNGKPNWRY-EPMGVCAGIASWNATFLYGVKMKIAPALAAAGCSFIFKASEKSLGV 190
DB 127 RTTPAGPGRFLNLTREPIGVVGAITPNFPAQVAVKVIAPALAMGNAIVLKPAQLAPLP 186
QY 191 LGLAPLFAEAGPPGVVOFLTG-ARVTEGALASHMDIAKISFTRSVGGGRAVKQATLKS 249
DB 187 VALGELAELAGLPGLVNLPGRSVAGNALVQHPFSGVTKFTGSTEVGQQIGRMA-ADR 245
QY 250 MKRVTLGEKPTIVFNEAPLERQSGESAKD-----PSKFGQIWWPPSCLLVO-- 297
DB 246 LITASLELGKKSALV-----AFGSSPKAAVAVFQAMYSNQGECTAPSLRLVERP 297
QY 298 -WGNLAE----KFGVHRHSGCGQRLWGNLPKRPHTGPFVDKSYQDRVLGNDVGD 352
DB 298 IYDEVVELVQARVEAARVG-----DPLDPTDTEIGPLISAEQRESVHVSVS 346
QY 353 TAQLLTGVGRKG-----DKGFAIEPTIFVNPKPGSKIWFEEIFGVLISIKFTKEEAIE 407
DB 347 EGATLISGGDSQPTGAPEGQFYRPTLFSGVYADMRAREEIGFVLSVLPFGECEAIT 406
QY 408 IANDTYGLASVIYKSLNRLGVSSALETGGVSINFPFIPETOTPEGGMKQSGREL 467
DB 407 LANDIVFGLAGVTRDVGRLRFAQTLDAGNWNWSGVLNPASPYRGFGQSGYSDLG 466
QY 468 EEGKAYLEPKTI 480
DB 467 QAAIESFTKRSI 479

RESULT 7
US-09-655-270A-9
; Sequence 9, Application US/09655270A
; Patent No. 6329151
; GENERAL INFORMATION:
; APPLICANT: KOUVIERE, PIERRE E
; FILE REFERENCE: BC1011 US NA
; CURRENT APPLICATION NUMBER: US/09/655,270A
; CURRENT FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/120,702

; Sequence 9, Application US/09651941
; Patent No. 6355470
; GENERAL INFORMATION:
; APPLICANT: ROUVIER, PIERRE E
; APPLICANT: WALTERS, DANA M
; APPLICANT: RAINIER, RUSS
; TITLE OF INVENTION: Genes Encoding Plicic Acid Degradation
; FILE REFERENCE: BC1022 US NA
; CURRENT APPLICATION NUMBER: US/09/651,941
; CURRENT FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,545
; PRIOR FILING DATE: 1999-10-03
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Rhodococcus erythropolis HL PM-1
; US-09-655-270A-9

Query Match      23.8%; Score 605.5; DB 4; Length 508;
Best Local Similarity 32.3%; Pred. No. 2.8e-59;
Matches 159; Conservative 81; Mismatches 206; Indels 47; Gaps 11;

QY 12 LFINFEVSSKGSERLTLTNPMDESTVATDVHVAADVDSDAASVAQVKKGPWKFTG 71
DB 33 LVIGDQLTPSSTGATFDSINPADGSHLAS-VAEATAADVARAVEAAKAAAR--TWQMRP 89
QY 72 AORACMLKPADLAEKNAEKLESPTGRPVSMITHFDIPNMVSVFRYYAGNADKIAG 131
DB 90 AORTLMFRYAALIEBEKTELAQSDMDGKPIRESLGIDLPIIMETLEYFAGLVTKIEG 149
QY 132 KTFPEDNGKPNWRY-EPMGVCAGIASWNATFLYGVKMKIAPALAAAGCSFIFKASEKSLGV 190
DB 150 RTTPAGPGRFLNLTREPIGVVGAITPNFPAQVAVKVIAPALAMGNAIVLKPAQLAPLP 209
QY 191 LGLAPLFAEAGPPGVVOFLTG-ARVTEGALASHMDIAKISFTRSVGGGRAVKQATLKS 249
DB 210 VALGELAELAGLPGLVNLPGRSVAGNALVQHPFSGVTKFTGSTEVGQQIGRMA-ADR 268
QY 250 MKRVTLGEKPTIVFNEAPLERQSGESAKD-----PSKFGQIWWPPSCLLVO-- 297
DB 269 LITASLELGKKSALV-----AFGSSPKAAVAVFQAMYSNQGECTAPSLRLVERP 320
QY 298 -WGNLAE----KFGVHRHSGCGQRLWGNLPKRPHTGPFVDKSYQDRVLGNDVGD 352
DB 321 IYDEVVELVQARVEAARVG-----DPLDPTDTEIGPLISAEQRESVHVSVS 369
QY 353 TAQLLTGVGRKG-----DKGFAIEPTIFVNPKPGSKIWFEEIFGVLISIKFTKEEAIE 407
DB 370 EGATLISGGDSQPTGAPEGQFYRPTLFSGVYADMRAREEIGFVLSVLPFGECEAIT 429
QY 408 IANDTYGLASVIYKSLNRLGVSSALETGGVSINFPFIPETOTPEGGMKQSGREL 467
DB 430 LANDIVFGLAGVTRDVGRLRFAQTLDAGNWNWSGVLNPASPYRGFGQSGYSDLG 489
QY 468 EEGKAYLEPKTI 480
DB 490 QAAIESFTKRSI 502

RESULT 8
US-09-346-611-2
; Sequence 2, Application US/08346611
; Patent No. 5684242
; GENERAL INFORMATION:
; APPLICANT: Schnable, Patrick S.
; APPLICANT: Wise, Roger P.
; TITLE OF INVENTION: NUCLEAR RESTORER GENES FOR HYBRID SEED
; TITLE OF INVENTION: PRODUCTION
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/346,611
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Silver, Donald J.
REGISTRATION NUMBER: 37552
REFERENCE/DOCKET NUMBER: 63632
TELEPHONE: (312) 616-5600
TELEFAX: (312) 616-5700
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 133 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-346-611-2

Query Match 9.1%; Score 231; DB 1: Length 133;
Best Local Similarity 43.7%; Pred. No. 5.1e-18;
Matches 52; Conservative 19; Mismatches 46; Indels 2; Gaps 2;
QY 162 LYVGWKIAPALAGCSIFFKASEKSPGLVGLAPLFAEAGPPGVVQFLTGARVT-GEAL 220
DB 4 LMYAKVGPALACGNTLVLTAEQTPLSXLYISKLLHEAGLPEGVYVNVVSGFGPTAGAA 63
QY 221 ASHMDIAKISFTRSGGAVKQATLKSNNKRVTLLEGEK-PTIVFNEAPLERQSGESA 278
DB 64 ASHMDVDKIAFTGTDGTGKILLEAAKSNLKTVTLELGGKSPFIWTKLMEGPAXGAXA 122

RESULT 9
US-08-794-494-2
Sequence 2, Application US/08794494
Patent No. 5981833
GENERAL INFORMATION:
APPLICANT: Wise, Roger P.
APPLICANT: Schnable, Patrick S.
TITLE OF INVENTION: NUCLEAR RESTORER GENES FOR HYBRID SEED PRODUCTION
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Leydig, Voit & Mayer, Ltd.
STREET: Two Prudential Plaza, Suite 4900
CITY: Chicago
STATE: IL
COUNTRY: US
ZIP: 60601-6780
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/794,494
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 8-346611
FILING DATE: 29-NOV-1994
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Larcher, Carol
REGISTRATION NUMBER: 35243
REFERENCE/DOCKET NUMBER: 74959
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 616-5600
TELEFAX: (312) 616-5700
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 133 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-794-494-2

Query Match 9.1%; Score 231; DB 2: Length 133;
Best Local Similarity 43.7%; Pred. No. 5.1e-18;
Matches 52; Conservative 19; Mismatches 46; Indels 2; Gaps 2;
QY 162 LYVGWKIAPALAGCSIFFKASEKSPGLVGLAPLFAEAGPPGVVQFLTGARVT-GEAL 220
DB 4 LMYAKVGPALACGNTLVLTAEQTPLSXLYISKLLHEAGLPEGVYVNVVSGFGPTAGAA 63
QY 221 ASHMDIAKISFTRSGGAVKQATLKSNNKRVTLLEGEK-PTIVFNEAPLERQSGESA 278
DB 64 ASHMDVDKIAFTGTDGTGKILLEAAKSNLKTVTLELGGKSPFIWTKLMEGPAXGAXA 122

RESULT 10
US-07-689-008-2
Sequence 2, Application US/07689008
Patent No. 5268274
GENERAL INFORMATION:
APPLICANT: Ben-Bassat, Arie
APPLICANT: Calhoon, Roger D
APPLICANT: Fear, Anna L
APPLICANT: Gelfand, David H
APPLICANT: Meade, James H
APPLICANT: Tai, Rony
APPLICANT: Wong, Hong
APPLICANT: Ben-Ziman, Moshe
TITLE OF INVENTION: METHODS AND NUCLEIC ACID SEQUENCES FOR THE
EXPRESSION OF CELLULOSE SYNTHASE OPERON
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: McCutchen, Doyle, Brown & Enersen
STREET: Three Embarcadero Center
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/689,008
FILING DATE: 19910422
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 337,194
FILING DATE: 12-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 496,236
FILING DATE: 23-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Murthy, Elisabeth Felix
REGISTRATION NUMBER: 31547
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 393-2000
TELEFAX: (415) 393-2286
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3031 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein

US-09-045-632-49

STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-045-632-50

Query Match 3.5%; Score 89.5; DB 3; Length 1050;

Best Local Similarity 20.1%; Pred. No. 1.6; Indels 137; Gaps 21;
Matches 72; Conservative 47; Mismatches 102;

QY 189 GVLGLAPLFAEAGPPGVQVFL-----TCARVTGALASHMDIAKISTRSVGGGRA 240
DB 701 GPLGITISGTEEPDPIIISGLTKRGLAERTGASMLGTA---YNPSKXSVSLK-----GRP 752
QY 241 VKQA--TLKSNMRKRVTLLECEKPTIVFNEAPLERQSG--ESAKDFSKFGQIWWPPSCL-- 294
DB 753 LSEAIHLQVAGETVTLKIKKQ---LDRPLLPROSGSLSEASVDVE-----DPPEALKG 803
QY 295 -----LVOMGNLAEKPHGVHRHSGFGGC-----ORMLGQN 323
DB 804 GLLTTHSPAVPSVDSAVESWGSSATE-----GGFGSGSYTPOVAVRSVTPQWRSSR 857
QY 324 -----PLEPKRTH---GP-----FVDKSOYD-----RVLGNI----- 347
DB 858 LKSSPPPLEPRRTSYTPGPTDESFPDEEGDWEPNPSAPGAREEGFWRVIGLEADLE 917
QY 348 DVGKD-----TAQLLTGVRKGDGKFAIEPTIFVNPXPGSKIW-----FEEI 389
DB 918 SCGOSSELLRELEASIMTGTQV-----SVAVDGRPGSRPMRRSRREVGTSPEDLQEL 967
QY 390 FGPVLSIKTKFTEEEAIEANDTGYGLASVIVTKSLN--RGLRVSSALETGGVSINFPF 446
DB 968 LLPT-PLEMHRVTLHKDPVRNDFGFSVSDGLEKGVYVHTVRIDGPAHQHGLQ---PF 1021

RESULT 13

US-09-045-632-50

Sequence 50, Application US/09045632

Patent No. 6001575

GENERAL INFORMATION:

APPLICANT: Huanair, Richard L.

APPLICANT: Dong, Hualing

TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND

TITLE OF INVENTION: GRIP-RELATED MOLECULES

NUMBER OF SEQUENCES: 105

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/045,632

FILING DATE: 19-MAR-1998

CLASSIFICATION: 435

Prior Application DATA:

APPLICATION NUMBER: US 60/041,016

FILING DATE: 19-MAR-1997

ATTORNEY/AGENT INFORMATION:

NAME: Corliss, Peter F.

REGISTRATION NUMBER: 33,860

REFERENCE/DOCKET NUMBER: 48147/1699-CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 50:

SEQUENCE CHARACTERISTICS:

LENGTH: 1050 amino acids

TYPE: amino acid

STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-045-632-50

Query Match 3.5%; Score 89.5; DB 3; Length 1050;

Best Local Similarity 20.1%; Pred. No. 1.6; Indels 137; Gaps 21;
Matches 72; Conservative 47; Mismatches 102;

QY 189 GVLGLAPLFAEAGPPGVQVFL-----TCARVTGALASHMDIAKISTRSVGGGRA 240
DB 701 GPLGITISGTEEPDPIIISGLTKRGLAERTGASMLGTA---YNPSKXSVSLK-----GRP 752
QY 241 VKQA--TLKSNMRKRVTLLECEKPTIVFNEAPLERQSG--ESAKDFSKFGQIWWPPSCL-- 294
DB 753 LSEAIHLQVAGETVTLKIKKQ---LDRPLLPROSGSLSEASVDVE-----DPPEALKG 803
QY 295 -----LVOMGNLAEKPHGVHRHSGFGGC-----ORMLGQN 323
DB 804 GLLTTHSPAVPSVDSAVESWGSSATE-----GGFGSGSYTPOVAVRSVTPQWRSSR 857
QY 324 -----PLEPKRTH---GP-----FVDKSOYD-----RVLGNI----- 347
DB 858 LKSSPPPLEPRRTSYTPGPTDESFPDEEGDWEPNPSAPGAREEGFWRVIGLEADLE 917
QY 348 DVGKD-----TAQLLTGVRKGDGKFAIEPTIFVNPXPGSKIW-----FEEI 389
DB 918 SCGOSSELLRELEASIMTGTQV-----SVAVDGRPGSRPMRRSRREVGTSPEDLQEL 967
QY 390 FGPVLSIKTKFTEEEAIEANDTGYGLASVIVTKSLN--RGLRVSSALETGGVSINFPF 446
DB 968 LLPT-PLEMHRVTLHKDPVRNDFGFSVSDGLEKGVYVHTVRIDGPAHQHGLQ---PF 1021

RESULT 14

US-09-503-172A-2

Sequence 2, Application US/09503172A

Patent No. 6284510

GENERAL INFORMATION:

APPLICANT: ITO, Tetsuya

APPLICANT: FUJITA, Koki

APPLICANT: HARA, Kozo

APPLICANT: TONOKURA, Takashi

APPLICANT: SAKANO, Yoshiyuki

TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE GENE

FILE REFERENCE: 10749-0001-0

CURRENT APPLICATION NUMBER: US/09/503,172A

CURRENT FILING DATE: 2000-02-14

Prior Application NUMBER: JP 160416/1999

Prior Filing DATE: 1999-06-08

NUMBER OF SEQ ID NOS: 9

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2

LENGTH: 578

TYPE: PRT

ORGANISM: Arthrobacter sp.

US-09-503-172A-2

Query Match 3.5%; Score 88.5; DB 4; Length 578;

Best Local Similarity 22.2%; Pred. No. 0.76;

Matches 121; Conservative 49; Mismatches 193; Indels 181; Gaps 28;

QY 20 SSKGSERLTLTNPDWDESTVATDVHANAADVDVAASVQAVKKGPWKKFTGAQRAACML 79

DB 4 STRGRRVRVLAGGLATSTLANAVLIAGAAP-----ATAQSGLODDGPEPTIHTQAYAPED 58

QY 80 KF-ADLAEKNAEKLALESPT--GRPVSMITHFDIPNKVSVF----- 119

DB 59 DFTAKWTRADARQLORM--SDPTAPSRNSPMASVTMTVPQDFPDMSNEQVWVWDWTPLT 117

QY 120 ----RYVA--GW-----ADKIAGTKTFEDNGKPNWRYPENMGVCAGIASWNAFTLYVGW 166

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:45:31 ; Search time 104.96 Seconds
(without alignments)
515.367 Million cell updates/sec

Title: US-09-882-694A-5
Perfect score: 2549
Sequence: 1 MVLSPDEYKSELFNNFEVFS.....EGLKAYLEPKTNIHWNIE 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/genesecp-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/genesecp-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/genesecp-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/genesecp-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/genesecp-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/genesecp-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/genesecp-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/genesecp-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/genesecp-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/genesecp-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/genesecp-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/genesecp-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/genesecp-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/genesecp-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/genesecp-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/genesecp-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/genesecp-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/genesecp-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/genesecp-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/genesecp-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/genesecp-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/genesecp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 2549 | 100.0 | 487 | 22 | AA1980 |
| 2 | 866 | 34.0 | 496 | 16 | AA1981 |
| 3 | 857 | 33.6 | 497 | 7 | AA1982 |
| 4 | 824.5 | 32.3 | 529 | 22 | AA1983 |
| 5 | 818.5 | 32.1 | 500 | 22 | AA1984 |
| 6 | 817.5 | 32.1 | 521 | 15 | AA1985 |
| 7 | 815.5 | 32.0 | 495 | 16 | AA1986 |
| 8 | 781.5 | 30.7 | 501 | 21 | AA1987 |
| 9 | 781.5 | 30.7 | 501 | 21 | AA1988 |
| 10 | 781 | 30.6 | 538 | 21 | AA1989 |
| 11 | 776.5 | 30.5 | 520 | 22 | AA1990 |

| | | | | | |
|----|-------|------|-----|----|----------|
| 12 | 772.5 | 30.3 | 659 | 22 | ABB62256 |
| 13 | 725 | 28.4 | 534 | 21 | AA1991 |
| 14 | 704 | 27.6 | 909 | 22 | ABB58519 |
| 15 | 700 | 27.5 | 424 | 21 | AA1992 |
| 16 | 684 | 26.8 | 495 | 22 | AA1993 |
| 17 | 681.5 | 26.7 | 497 | 21 | AA1994 |
| 18 | 681.5 | 26.7 | 501 | 21 | AA1995 |
| 19 | 666.5 | 26.1 | 496 | 22 | AA1996 |
| 20 | 665 | 26.1 | 490 | 22 | AA1997 |
| 21 | 660.5 | 25.9 | 496 | 22 | AA1998 |
| 22 | 660.5 | 25.9 | 496 | 22 | AA1999 |
| 23 | 655.5 | 25.7 | 503 | 21 | AA2000 |
| 24 | 655.5 | 25.7 | 503 | 21 | AA2001 |
| 25 | 645 | 25.3 | 490 | 22 | AA2002 |
| 26 | 629.5 | 24.7 | 477 | 22 | AA2003 |
| 27 | 619.5 | 24.3 | 508 | 18 | AA2004 |
| 28 | 612.5 | 24.0 | 497 | 15 | AA2005 |
| 29 | 612.5 | 24.0 | 505 | 17 | AA2006 |
| 30 | 606.5 | 23.8 | 481 | 22 | AA2007 |
| 31 | 605.5 | 23.7 | 508 | 21 | AA2008 |
| 32 | 604.5 | 23.7 | 507 | 22 | AA2009 |
| 33 | 604.5 | 23.7 | 507 | 22 | AA2010 |
| 34 | 601.5 | 23.6 | 497 | 16 | AA2011 |
| 35 | 599.5 | 23.5 | 474 | 22 | AA2012 |
| 36 | 587 | 23.0 | 357 | 21 | AA2013 |
| 37 | 582.5 | 22.9 | 497 | 22 | AA2014 |
| 38 | 582.5 | 22.9 | 497 | 22 | AA2015 |
| 39 | 578.5 | 22.7 | 512 | 22 | AA2016 |
| 40 | 578.5 | 22.7 | 512 | 22 | AA2017 |
| 41 | 575 | 22.6 | 496 | 22 | AA2018 |
| 42 | 572 | 22.4 | 391 | 21 | AA2019 |
| 43 | 569 | 22.3 | 314 | 21 | AA2020 |
| 44 | 566 | 22.2 | 393 | 21 | AA2021 |
| 45 | 566 | 22.2 | 393 | 21 | AA2022 |

ALIGNMENTS

RESULT 1

AA1980

ID AA1980 standard; Protein; 487 AA.

AC AA1980;

XX 02-MAY-2001 (first entry)

DT Exophiala spinifera aldehyde dehydrogenase.

DE Fumonisin; degradative enzyme; transport enzyme; flavin monooxygenase;

KW aldehyde dehydrogenase; permease; p-glycoprotein; fumonisin esterase;

KW detoxification; mycotoxin; animal feed; human feed; silage;

KW transgenic plant; transgenic animal; microbial spray.

XX Exophiala spinifera.

OS WO200105980-A1.

XX 25-JAN-2001.

XX 14-JUL-1999; 99WO-US15824.

XX 14-JUL-1999; 99WO-US15824.

PR (PION-) PIONEER HI-BRED INT INC.

PA (CURA-) CURAGEN CORP.

XX Duvick JP, Maddox J, Gilliam J, Folkerts O, Crasta OR;

PI WPI; 2001-147345/15.

DR N-PSDB; AAD02692.

XX Novel polynucleotides encoding Exophiala degradative or transport

PT

enzyme which is useful for detoxifying fumonisin or structurally related mycotoxin during processing of grain for human or animal food consumption -

Claim 1c; Page 68-70; 90pp; English.

The patent discloses novel polynucleotides encoding *Exophiala spinifera* fumonisin degradative or transport enzymes such as flavin monooxygenase, aldehyde dehydrogenase, permease and p-glycoprotein. A fumonisin esterase enzyme with at least one fumonisin degradative enzyme is useful for detoxifying fumonisin or a structurally related mycotoxin present in harvested grain, on application to a plant or to harvested grain during processing, or to processed grain that is to be used as animal or human feed, or as a silage. Nucleotide fragments of the present invention are useful as probes and primers. They can be introduced into microorganisms that multiply on plants to deliver enzymes to potential target crops. The genes encoding the degrading enzymes are introduced via a vector into a microbial host and the transformed host is supplied to the environment, plants or animals for reducing the pathogenicity of a fungus producing fumonisin. The genes of the invention are fermented in a bacterial host and the resulting bacteria is processed and used as a microbial spray. The nucleotide sequences can be used alone or in combination to engineer microbes or other organisms to metabolise fumonisin and resist its toxic effects. The present protein sequence is aldehyde dehydrogenase, a fumonisin degradative enzyme from *Exophiala spinifera*.

Sequence 487 AA;

```

Query Match      100.0%; Score 2549; DB 22; Length 487;
Best Local Similarity 100.0%; Pred. No. 2.5e-237;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVLSPDEYKSELFINNEFYSSKGSERLTITNPWDESTVATDVHANAADVDSAVAASVQA 60
DB 1 mvlspdeykselfinnefysskgserltitnpwdestvatdvhanaadvsavaasvqa 60

QY 61 VKKGPWKKTGTQRAACMLKFDALAEKNAEKLARLESPTGRPVSMITHFDIPNMVSVFR 120
DB 61 vkkgpwkktgtqraacmlkfdalaecknaeklarlesptgrpvsmtihfdipnmvsvfr 120

QY 121 YTAGWADKTAGTTFPEDNGKPNWRYEPPMGVCAGIASWNATFLYVGWKIAPALAAAGCSFIF 180
DB 121 ytagwadkiagtktfpedngkpnwryepmgvcagiaswnatflvygwkiaapalaagcsfif 180

QY 181 KASEKSPGLGVLGLAPLFAEAGPPGVVORLTGARVTGEALASHMDIAKTSFTRSVGGRA 240
DB 181 kasekspglgvlglaplfaeagppgvvqrltgartvtegealashmdiaktsftrsavggra 240

QY 241 VKQATLKSNMKRVTLLEGEKPTIVFEAPLERQSGESAKDFSKFGQITWPPSCLLYQWGN 300
DB 241 vkqatlksnmkrvtllegekptivfeaplerqsgesakdfskfgitwppscillyqwn 300

QY 301 LAEKTHGVHRHSGCGCORWLGQNPLEPKRTHGPFVDKSDYDRVLGNIDVCKDPAQLLTGV 360
DB 301 laekthgvhrhsgcgcorwlgqnplepkrthgpfvdksdydrvlgnidvckdpaqltgv 360

QY 361 GRKGDKGFAIEPTIFVNPKGSKWFEELFGPVLSTIKTKTEEEAIEIANDITYGLASVI 420
DB 361 grkgdkgfaieptifvnpkgskwfeelfgpvlstiktkteeeaielandityglasvi 420

QY 421 YTKSLNRLGRVSSALETGVSINFPFIPETQTPFGGMKQSGSGRELGEGLKAYLPKTI 480
DB 421 ytkslnrlgrvssaletgvsinfpfpetqtpfggmksqsgsrelgeglkaylpkkti 480

QY 481 NIHVNIE 487
DB 481 nihvnie 487

```

RESULT 2
AAR71891

```

ID AAR71891 standard; Protein; 496 AA.
AC AAR71891;
XX
XX 25-OCT-1995 (first entry)
XX
DE Cladosporium herbarum allergen Clah53.
XX
XX Fungal spore; allergen; Clah53; allergy; aldehyde dehydrogenase.
XX
OS Cladosporium herbarum.
XX
XX W09506121-A.
XX
XX 02-MAR-1995.
XX
XX 24-AUG-1994; 94WO-AT00120.
XX
XX 27-AUG-1993; 93AT-0001725.
XX
XX (BIOM-) BIOMAY PRODN & HANDELSGES MBH.
XX
XX Achatz G, Breitenbach M, Ebner C, Hirschwehr R;
XX Kraft D, Lechenauer E, Oberkofler H, Prillinger H;
XX Simon B, Unger A;
XX
XX WPI: 1995-106850/14.
XX N-PSDB; AAQ86278.
XX
XX Allergens derived from Cladosporium herbarum spores - also
XX recombinant DNA for expressing the allergens, useful for in vitro
XX allergy detection
XX
XX Claim 1; Page 8-9; 35pp; German.
XX
XX Spores of Cladosporium herbarum are the most common fungal spores
XX found in the air; they can cause allergic reactions. Various Clah
XX allergens and sequences encoding them have now been isolated. The
XX mature Clah53 allergen has mol. wt. 53 kD and is encoded by cDNA
XX sequence AAQ86278. The allergen has homology to aldehyde
XX dehydrogenases. Potential epitopic subfragments were identified by
XX computer analysis of the amino acid sequence. See AAR71892-R71906 for
XX potential B-cell epitopes and AAR72615-R72627 for potential T-cell
XX epitopes.
XX
XX Sequence 496 AA;

Query Match      34.0%; Score 866; DB 16; Length 496;
Best Local Similarity 42.8%; Pred. No. 9.7e-75;
Matches 211; Conservative 67; Mismatches 183; Indels 32; Gaps 12;

QY 7 EYKSELFINNEFYSSKGSERLTITNPWDESTVATDVHANAADVDSAVAASVQAVKKGPW 66
DB 15 eqptglfinnefysskgsertitnpwdestvatdvhnaadvsavaasvqavkkgpw 66

QY 67 KKFTGAORAAACMLKFDALAEKNAEKLARLESPTGRPVSMITHFDIPNMVSV----PRY 121
DB 73 rletpenrgklinnlanlfeantdlaaevsidngkatsm-----arvtasacagclry 126

QY 122 YAGWADKTAGTTFPEDNGKPNWRY---EPMGVCAIASWNATFLYVGWKIAPALAAAGCSF 178
DB 127 yggwadkitgkvi--dctpdctfnvykkepigncrsdhsalepllmwawkiapalacntv 184

QY 179 IFKASEKSPGLGVLGLAPLFAEAGPPGVVQPLTG-ARVTGEALASHMDIAKTSFTRSVGG 237
DB 185 vlktaeqtpigglvaaslvkeagfppgvvinvisgfvagaaalshmdvdkvafgtgsv 244

QY 238 GRAVKQATLKSNMKRVTLLEGEK-PTIVFEAPLERQ-SGESAKDFSKFGQITWPPSCLL 295
DB 245 grtllkaasnsnlkvltlelggkspnlvfedaidlnalsvwnfglffhngqccagary 304

QY 296 VQ---WGNLAEKPHGVHRHSGCGCORWLGQNPLEPKRTHGPFVDKSDYDRVLGNIDVGKD 352

```

Db 305 vqesi yd k f v g k f k -----eragknvvgdbfaadtfggqvskvqfdrimey l q a g k d 357
 QY 353 T-AQLLTGVRGKDGKGFALTEPTIFVNPXPGSKWFEETFGPVLSIKTKPTEEEAIEIAND 411
 Db 358 agatvetsgrkdkykfyieptifsnvtedmkivkeelfgpvcslakfktdaiki g n a 417
 QY 412 TTYGLASVYTKSLNGLRVSSALETGGVSNFPFIPETQTPFGGKMGKSGSGRELGEGL 471
 Db 418 styglaaavhtknlncaievsnalkagtwwntyntnlhhqmpfggykesgigrelgedal 477
 QY 472 KAYLEPKTINIHV 484
 Db 478 anytqktvsirl 490

RESULT 3
 AAP60456
 ID AAP60456 standard; Protein; 497 AA.

XX AAP60456;
 AC
 DT 25-JUN-1991 (first entry)
 XX

DE Sequence of aldehyde dehydrogenase (aldh) of *Aspergillus nidulans*.
 XX
 KW Filamentous fungi promoter.
 XX
 OS *Aspergillus nidulans*.
 XX
 PN W08606097-A.
 XX
 PD 23-OCT-1986.

XX 14-APR-1986; 86WO-GB00209.
 XX
 PR 20-DEC-1985; 85US-0811404.
 XX
 PR 15-APR-1985; 85CA-0479135.
 XX
 PA (ALLE-) ALLELIX INC.
 XX
 PI Gwynne DL, Buxton F, Pickett M, Davies R, Scazzocchio C;
 XX
 DR WPI; 1986-291664/44.
 DR
 DR N-PSDB; AAN60401.

XX DNA construct for use in filamentous fungi - comprising promoter
 PT operative in filamentous fungi to promote transcription of coding
 PT region
 PT
 PS Disclosure; Fig 1A; 75pp; English.

XX In the constructs of the invention, the promoter region naturally
 CC associated with the alcohol dehydrogenase I (aldh) gene and the
 CC aldehyde dehydrogenase (aldh) gene of *A. nidulans* or naturally
 CC associated with the glucosylase gene in *Aspergillus niger* may be
 CC used. The DNA construct may contain a promoter region in operative
 CC association with a signal peptide coding region. The promoter/signal
 CC construct is suitably provided with a flanking restriction site to
 CC allow precise coupling of the protein coding region to the signal
 CC peptide coding region.

XX Sequence 497 AA;

Query Match 33.68; Score 857; DB 7; Length 497;
 Best Local Similarity 42.1%; Pred. No. 7.2e-74;
 Matches 202; Conservative 79; Mismatches 177; Indels 22; Gaps 11;

QY 12 LFTNNEFVSKSGSERLTLPNWDSTVATDVHVANAADVDSAAVAVKRGKPKFTG 71
 Db 21 Lfinnefvkgvektqgvinsnek-vitsvbeatekdvdvavaaa-raafegpqrqvtp 78

QY 72 AGRACMLKFAADLAENAEKLESLPTGPRVSMITHEDIPNMVSVRYAGNADKIAG 131
 Db 79 sergilinkladmeriditlaaiesldngkaftm-akvdlansigclryyagwadkihg 137
 QY 132 KTFPEDNGKPNW-RYEDMGVCAGIASWNTATLYVGWKTAPALAAAGCSFIPKASEKSPGV 190
 Db 138 qtdtapteltlyrhepvgvcgqilpwnfpilmwsklgpavaagntvvlktadqtplsa 197
 QY 191 LGLAPLFAAGPPGVGVFLTG-ARVTCGEALASHMDIAKISFTRSVGGGRAVKOATUKSN 249
 Db 198 lyaaklikeapfpagvinvisgfrtagaaishmdickvaftgstlvpgptilqaaaksn 257
 QY 250 MKRVTLLEGEK-PTIVFNEAPLERQ-SGESAKDFSGQIWWPPSCLLVQWGNLAERF-- 305
 Db 258 lkkvtllelggkspnivfdadidnaishwanfigfnhgqccagrsrlivqeg-iydhfva 316
 QY 306 ---HGVRHSGFGGQRMWLGQNPLEPKRTHGPFVDKQSDRVLGNIDVCKDT-AOLLTGVG 361
 Db 317 rfkeragknkv-----npfeqdtfggbqvsqglgfdriimeyinhgkagatvatggd 368
 QY 362 RXGDKGFAJTEPTIFVNPXPGSKWFEETFGPVLSIKTKPTEEEAIEIANDTTYGLASVIY 421
 Db 369 rhgneyfiqptvftdvtmdmkiagceifgpvvtikfkdvaaekikgstydlaaavh 428
 QY 422 TKSLLNGLRVSSALETGGVSNFPFIPETQTPFGGKMGKSGSGRELGEGLKAYLEPKTIN 481
 Db 429 tknvtairvsnalkagtwwlnnymisyaqpfqgkgsglgrlgelgsyalenytqiktvh 488

RESULT 4
 ABG06577
 ID ABG06577 standard; Protein; 529 AA.

XX ABG06577;
 AC
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #6568.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX

OS Homo sapiens.

XX W0200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS70764.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

XX Claim 20; SEQ ID No 36936; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques

to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG0377 represent novel human diagnostic amino acid sequences. The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 529 AA;

Query Match 32.3%; Score 824.5; DB 22; Length 529;
Best Local Similarity 41.1%; Pred. No. 1.1e-70;
Matches 198; Conservative 78; Mismatches 189; Indels 17; Gaps 11;

QY 10 SELFINNEFVSSKGSRLTITNPWDESVATDVHANAADVDVAASVQAVKKG-PKK 68
Db 49 tkifinnewhesksgkfctncpstreic-eeegdkpvdvkaaveaaqvafrgspwrr 107
QY 69 FTGAQRAACMLKPADLAENAKELARLESPTGRPVSMITHFDIPNMYSVFRYAGWADK 128
Db 108 ldaIsgrllhqladlverdratlaalemdtgkpfllhaffldlegcirtlryagwadk 167
QY 129 IAGTTP-EDNGKPNRYPMPGVCAGIASNNATFLYVGWKTAPALAACCSFIFKASEKSP 187
Db 168 lqgktpidavvcfrthepigvcgaitpwnfpllmvwxklapalccntmvlkpaecqp 227
QY 188 LGVUGLAPLFAEAGFPVGVQFLTGARVT-GEALASHMDIAKISTRSVGGGRAVKQATL 246
Db 228 ltalylqslkiesagfpggvvniyvgftvgaaishbpqinkiaftgstevgklvkeas 287
QY 247 KSNMKRVTLGLGK-PTIVFNEAPLERQGESAKD--FSKGIQVVPVSCLLVQWGNLAE 303
Db 288 rsnlkrvtlelgknpccivcadadld-lavecaghvfnqgcqtaasrfve----e 341
QY 304 KFHG--VRHSGFGCGRWLQGNPLEPKRTHGPFVDKSDYVLGNIDVG-KDTAQLLTGV 360
Db 342 qvysefvrrsveyakrpvg-dpfdvkteqgpdqdkgfdkilleiesgkkekalecgg 400
QY 361 GRKDGKGAIEPTIFVNPXPGSKIMPEEIFGPVLSIKTFKTEEAIEIANDTTGLASVI 420
Db 401 samedkgllfkpcvfsevtndmrakceefgvpvqpllkfksteevikranstdygltaav 460
QY 421 YTKSLNRGLRVSSALETGGVSNFPPIPETOTFGGMKSGSGRELGEGLKAYLEPKTI 480
Db 461 ftknldkalkiasalesgtwincynalyaqapfggfkmsngnreigeiyaleteykv 520
QY 481 NI 482
Db 521 ti 522

RESULT 5

AA874924
ID AAB74924 standard; Protein; 500 AA.
XX
AC AAB74924;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human aldehyde dehydrogenase ALDH2 protein sequence SEQ ID NO:4.
XX
KW Aldehyde dehydrogenase; glycerol dehydratase; 3-HP; glycerol;
KW feedstock; 3-hydroxypropionic acid; genetic engineering; glucose;
KW bacterial host; absorbable prosthetic device; surgical suture;

KW beta-lactam; acrylic acid; trifluoromethylated alcohol; diol;
KW polyhydroxyalkonate; copolymer; lactic acid.
XX
OS Homo sapiens.
XX WO200116346-A1.
XX
PD 08-MAR-2001.
XX
PF 30-AUG-2000; 2000WO-US23878.
XX
PR 30-AUG-1999; 99US-0151440.
XX
PA (WISC) WISCONSIN ALUMNI RES FOUND.
XX
PI Suthers PF, Cameron DC;
XX
XX WPI; 2001-315988/33,
DR N-PSDB; AAF82082.
DR
XX 3-Hydroxypropionic acid preparation, for use e.g. as monomer, by
PT fermenting recombinant microorganisms expressing genes for suitable
PT enzymes in the presence of glycerol or glucose
XX
PS Claim 5; Page 32-35; 63pp; English.
XX
CC The present invention describes a method for the production of
CC 3-hydroxypropionic acid (3-HP). The method comprises fermenting a
CC recombinant microorganism in the presence of a source of glycerol (I)
CC or glucose, where the microorganism: (i) expresses genes for non-native
CC enzymes which catalyse the production of (3-HP) from (I); (ii) carries
CC genetic constructions for the expression of a glycerol dehydratase
CC (GDR) and aldehyde dehydrogenase (ADH) capable of catalysing the
CC production of (3-HP) from (I); or (iii) carries a genetic construct
CC which expresses the dhap gene from Klebsiella pneumoniae and a gene for
CC an ADH capable of catalysing the production of (3-HP) from (I). 3-HP is
CC a monomer, and is useful e.g. in the production of absorbable prosthetic
CC devices and surgical sutures or for incorporation into beta-lactams,
CC production of acrylic acid or formation of trifluoromethylated alcohols
CC or diols, polyhydroxyalkonates and copolymers with lactic acid.
CC Incorporation of genes encoding two enzymes makes the host organisms
CC able to produce (3-HP) from (I). The biotechnological method of
CC preparing (3-HP) is potentially cheaper than chemical synthesis. The
CC present sequence represents the human aldehyde dehydrogenase ALDH2, which
CC is used in the exemplification of the present invention.
XX
SQ Sequence 500 AA;

Query Match 32.1%; Score 818.5; DB 22; Length 500;
Best Local Similarity 40.4%; Pred. No. 3.8e-70;
Matches 197; Conservative 85; Mismatches 191; Indels 15; Gaps 12;
QY 5 PDYKSELFINNEFVSSKGSRLTITNPWDESVATDVHANAADVDVAASVQAVKKG 64
Db 16 pevfcngqfinnewhdavarkftvpnp-stgevicqvaegdkedvdkaregrpgafglg 74
QY 65 -PMKFTGAQRAACMLKPADLAENAKELARLESPTGRPVSMITHFDIPNMYSVFRYIA 123
Db 75 spwrmdashsgrllnrliadlierdrtylaletidngkpyvisylvldmvlkrlrya 134
QY 124 GWADKIAGTTFPEDNGKPNW-RYEPWGVCAAGIASNNATFLYVGWKTAPALAACCSFIFKA 182
Db 135 gwadkyhgktlpidldgdfayctrhepgvcgqllpwnfpllmqawklgpalatgtnvnmkv 194
QY 183 SEKSPGLVGLAPLFAEAGFPVGVQFLTGARVT-GEALASHMDIAKISTRSVGGGRAV 241
Db 195 acqtpitalyvanlikeagfpggvvniyvgftvgaaishbedvkvafgstelgrvl 254
QY 242 KQATLKSNNKRVTLGLGK-PTIVFNEAPLERQGESAKDFSKP---GQIVVPVSCLLVQ 297
Db 255 gvaagsnkrvtlelggkspnlmsdadmd-waveqa-hraifndgqccaggrtfvq 312

| | | |
|----------|-----------|--|
| RESULT | 8 | |
| AAG36239 | | |
| ID | AAAG36239 | standard; Protein; 501 AA. |
| XX | XX | |
| XX | AC | AAG36239; |
| XX | XX | |
| DT | DT | |
| XX | XX | 18-OCT-2000 (first entry) |
| DE | DE | Arabidopsis thaliana protein fragment SEQ ID NO: 44383. |
| XX | XX | |
| XX | XX | Protein identification; signal transduction pathway; metabolic pathway; |
| KW | KW | hybridisation assay; genetic mapping; gene expression control; promoter; |
| XX | XX | termination sequence. |
| XX | XX | |
| OS | OS | Arabidopsis thaliana. |
| XX | XX | |
| PN | PN | EP1033405-A2. |
| XX | XX | |
| PD | PD | 06-SEP-2000. |
| XX | XX | |
| PF | PF | 25-FEB-2000; 2000EP-0301439. |

XX The invention provides nucleic acids encoding Arabidopsis plastidic
 CC acetyl CoA synthetase (ACS), various subunits (specifically the E3
 CC subunit) of plant plastidic pyruvate dehydrogenase (pPDH), the A and B
 CC subunits of plant ATP citrate lyase (ACL), Arabidopsis pyruvate
 CC decarboxylase (PDC), Arabidopsis aldehyde dehydrogenase (ALDH),
 CC specifically ALDH-2 and ALDH-4. The polypeptides can be expressed by
 CC standard recombinant methodology. The ACS, pPDH, ACL, PDC and ALDH
 CC polypeptides, methods and nucleic acid molecules of the invention are
 CC used to alter the level of acetyl CoA in a plant or plant cell, tissues
 CC or organs. A decrease in acetyl CoA is expected to affect the
 CC biosynthesis of very long chain fatty acids and flavonoids. The enzymes
 CC may also be used for in vitro synthesis of acetyl CoA, which in turn can
 CC be used to produce acetyl CoA phytochemicals (plastidic ACS, pPDH, ACL,
 CC pyruvate decarboxylase, acetyl CoA hydrolase, mitochondrial pyruvate
 CC dehydrogenase and aldehyde dehydrogenase are involved in the biosynthesis
 CC of acetyl CoA.

XX Sequence 538 AA;

Query Match 30.6%; Score 781; DB 21; Length 538;
 Best Local Similarity 38.6%; Pred. No. 1.8e-66;
 Matches 188; Conservative 81; Mismatches 196; Indels 22; Gaps 11;

Qy 10 SELFNNFVSSKSGSERLTLPNWDSTVATDVHVA--NAADVSAVAASVQAVKGPWK 67
 Db 58 tqllngfnvdasgkftfldprtgevia--hvaeagdaednrvakartafdegpwp 114
 Qy 68 KFTGAQAAACMLKFAADLAENAKELARLESPTGRPVSMITHFDIPNMVSVFYAGWAD 127
 Db 115 kmsayersrvllrfadvlkehseelasletwngkpyqsltaelpmfarliryyagwad 174
 Qy 128 KIAGKTPPED-NQKPNWRYEPPMGVCAGIASWNATFLYVGWKTAPALAGCSFIFKASEKS 186
 Db 175 kihgltipadgnyqvhthepigvaggilpwnfpllmfawkvgpalacgntivlvttaeqt 234
 Qy 187 PLGVGLAPLFAEAGPPGVQVFLTGART-GEALASHMDIAKISFTRSVGGGRAVKQAT 245
 Db 235 pltafyagkifleaglpvgvlnivsgfgatagaalashmdvdklaftgtdgtgkvilgla 294
 Qy 246 LKSNMKRVLTLEGEK-PTIVFNAPLERQSGESAKDFSKFGQIGVPPSCLLVQWGN--L 301
 Db 295 ansnlkptvltelgkspvifedadidk-----avealahalfnqgqccca--gsrtfv 347
 Qy 302 AEKPHG--VRHSGFGGQORWLGONPLEKPKRTHGPFVDKSOYDRVLGNIDVG-KDTAQLLT 358
 Db 348 hekvydefekstaralkrvvg-dprkgleggpqldlkqfekvmkyksgiesnatleec 406
 Qy 359 GVRKRGDKGPALEPTIFVNPKPQSGKIWFEEIFGPVLSIKTFKTEEPAIEIANDTTYGLAS 418
 Db 407 ggdqigdkgyfiqptvfnvkdmliaqdeifgpvqslkfsdvdevikranetkyglaa 466
 Qy 419 VIYTKSLNRLGVSSALETGVSNINFPFIPETOTPGGKQSGSGELGEEKLXLEYEPK 478
 Db 467 gvfknltdanrsralakgtvwnvcfdvdaaipfgygmksngnrekgylslnhnyldik 526
 Qy 479 TINHVN 485
 Db 527 avvtaln 533

RESULT 11

ID ABB60140 standard; Protein; 520 AA.

XX ABB60140;

AC ABB60140;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 7212.

DE Drosophila; developmental biology; cell signalling; insecticide;

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.
 XX Drosophila melanogaster.
 OS WO200171042-A2.
 PN 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US09231.
 PF 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 PA (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI: 2001-656860/75.
 DR N-PSDB; ABL04243.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PT
 XX Disclosure; SEQ ID NO 7212; 2lpp + Sequence Listing; English.
 PS
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 520 AA;

Query Match 30.5%; Score 776.5; DB 22; Length 520;
 Best Local Similarity 37.5%; Pred. No. 4.6e-66;
 Matches 183; Conservative 91; Mismatches 201; Indels 13; Gaps 11;

Qy 4 SPDEYKSELFTNNFEVSSKSGSERLTLPNWDSTVATDVHVAADVDASAASVQAVKK 63
 Db 34 tpdilytgvinnewhkskskglfetiinpttaevia-eiqcadkdidlavqarnafkl 92
 Qy 64 G-PWKKFTGAQRAACMLKFADLAENAKELARLESPTGRPVSMITHFDIPNMVSVFY 122
 Db 93 gspvrmdasergllyrladlmerdqvyasletidngkpsmsynovdptaiknlryf 152
 Qy 123 AGWADKTAGKTFPEDNCKPNW-RYEPNGVCAGIASWNATFLYVGWKTAPALAGCSFIFK 181
 Db 153 agwadknhgktfmdgofftyrthepvgvcgqilpwnfpilmmawklgpalatgnctivilk 212
 Qy 182 ASEKSPGLVGLIAPLFAEAGPPGVQVFLTGARTVGEALASHMDIAKISFTRSVGGGRAV 241
 Db 213 paeqtsltalyiaqlvkeagfpegvnnvpgfataalanhcdvkvafgtgtdvgkli 272
 Qy 242 KQATLKSNMKRVLTLEGEK-PTIVFNAPLERQSGESAKD--FSKFGQIHWPPSCLLVQW 298
 Db 273 qlasgntnlkrvtlglgkspnllstdmd-yavetahfglfnmgqcccgartfve- 330
 Qy 299 GNLAEEKFHGVRHSGFGGQORWLGONPLEKPKRTHGPFVDKSOYDRVLGNIDVG-KDTAQLL 357
 Db 331 dkiydef--versaerakktvg-npfdlnteqgppvneeqmeklmgiktgkqgaki 387
 Qy 358 TGVGR-KGDGFALIEPTIFVNPKPQSGKIWFEEIFGPVLSIKTFKTEEPAIEIANDTTYGL 416
 Db 388 aggsrpeglpgyfvgtvfadvqddmtiaraeelifgpvgqlirfkiddevierannseygl 447

CC subunits of a plant ATP citrate lyase (ACL), Arabidopsis pyruvate
CC decarboxylase (PDC), Arabidopsis aldehyde dehydrogenase (ALDH),
CC specifically ALDH-2 and ALDH-4. The polypeptides can be expressed by
CC standard recombinant methodology. The ACS, pPDH, ACL, PDC and ALDH
CC polypeptides, methods and nucleic acid molecules of the invention are
CC used to alter the level of acetyl CoA in a plant or plant cell, tissues
CC or organs. A decrease in acetyl CoA is expected to affect the
CC biosynthesis of very long chain fatty acids and flavonoids. The enzymes
CC may also be used for in vitro synthesis of acetyl CoA which in turn can
CC be used to produce acetyl CoA phytochemicals (plasticidic ACS, pPDH, ACL,
CC pyruvate decarboxylase, acetyl CoA hydrolase, mitochondrial pyruvate
CC dehydrogenase and aldehyde dehydrogenase are involved in the biosynthesis
CC of acetyl CoA.
XX
SQ Sequence 534 AA;

Query Match 28.4%; Score 725; DB 21; Length 534;
Best Local Similarity 37.2%; Pred. No. 4.5e-61;
Matches 181; Conservative 82; Mismatches 193; Indels 30; Gaps 12;

QY 10 SELFNNFEVSSKSERLTLTNPWDESTVATDVHVAADVDGSAVASQVAVKKGPKKF 69
DB 54 tqlllggrfdavsgkftptldprngevia-qvsegdaedvnravaarkatdegwpkm 112
QY 70 TGAORACMLKFAADLAENAEKLALESPTGRPVSMITHFDIPNMVSVFRYAGWADKI 129
DB 113 tayerskilfradliekhndetaaletwngkpyeqsaqlvypmlarvryyagwadki 172
QY 130 AGKTFPEDNGKPN---WRYEPMGVACIASWNNATFLYVGWKIAPALAAAGCSFIFKASEKS 186
DB 173 hgmtmpgdg--phivqtlhepigvaglipwnfplmlskwklpalacntvvlktaeqt 230
QY 187 PLGVGLIAPLPAEAGPPGVVQFLTGARVT-GEALASHMDIAKISFTRSVGGGRVAKQAT 245
DB 231 plsalllvgkllhesglpdygvnviwsgfgatagaiaashmdvkvafgstgdvgkillela 290
QY 246 LKSNMKRVTLGLGK-PTIVFNEAPLERQSGESAKDFSKF---GOIWVPPSCLLVQ---W 298
DB 291 sksnkavtlleeshsfvcedadvd-qavela-hfalifngdgqccagstfvhervy 348
QY 299 ONLAEKHGVRHGSFGGQCORWLGONPLEPKRT---HGPFVDKSYQYDRVLGNIDVGDKT-A 354
DB 349 defvekaka-----ralkrnvgdptfkslglegppqvdsqfknkilkylkhgveaga 398
QY 355 QLLTGVGKGDGKFAIPTFVNPCKGSKIWFREITGCVLSIKTEKTEEAIEIANDTNY 414
DB 399 tlqsggddc19skgyyigptvtsdvkddmlatdeifgqvctilkfkdleviaraunsty 458
QY 415 GLASVIYTKSLNRLGVSSALETGGVSINFPFIPETQTPFGMKQSGSGRELGEGLKAY 474
DB 459 glaagvftqltdahrimralrvgtvwinfcvldasipfgykmsgigregkgyislnny 518
QY 475 LEPKTI 480
DB 519 lqvKav 524

RESULT 14
ID ABB58519
XX ID ABB58519 standard; Protein; 909 AA.
AC ABB58519;
XX
DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 2349.
DE Drosophila; developmental biology; cell signalling; insecticide;
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX

PN W0200171042-A2.
PD 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL02622.
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 2349; 21pp + Sequence Listing; English.
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 909 AA;

Query Match 27.6%; Score 704; DB 22; Length 909;
Best Local Similarity 34.7%; Pred. No. 1.1e-58;
Matches 172; Conservative 89; Mismatches 196; Indels 38; Gaps 13;
QY 10 SELFNNFEVSSKSERLTLTNPWDESTVATDVHVAADVDGSAVASQVAVKKGPKKF 69
DB 429 tqllfingefvdaeqrtleivnptnee-vlckvacasatdvkavraahsaf-ygsvrq 486
QY 70 TGAORACMLKFAADLAENAEKLALESPTGRPVSMITHFDIPNMVSVFRYAGWADKI 129
DB 487 tprgrgqlmlnldlmernekeleatesvsgavvytlalkthvgmsieavryfagwcdkl 546
QY 130 AGKTFPEDNGKPN---WRYEPMGVACIASWNNATFLYVGWKIAPALAAAGCSFIFKASE 184
DB 547 ggntipvnpapnvnltfttrkepigvcgltcpwnypmlmswknaciasagntclikpaq 606
QY 185 KSPLGVGLIAPLPAEAGPPGVVQFLTG-ARVTGEALASHMDIAKISFTRSVGGGRVAKQ 243
DB 607 tcpltalfaeltvragfpvgvinvlpqgsdagvadhelvfkigtstipgkhmk 666
QY 244 ATLKSNMKRVTLGLGK-PTIVFNEAPLER--QSGESAKDFSKFQIOWVPPSCLLVQGN 300
DB 667 scadsnlkkslelqgkspilifadcdmdkavkhgmsvffnk-genciaagrifve--- 722
QY 301 LAERFH-----GVRHGSFGGQCORWLGONPLEPKRTGCPFVDKSOYDRVLGNIDVGK 351
DB 723 --drihdefirrvlkdrlmtig-----dpidrstangpqhkanfkilefcrgv 772
QY 352 DT-AOLLTGVGKGDG-KGFAIEPTFVNPCKGSKIWFREITGCVLSIKTEKTE--EBAE 407
DB 773 degaklyvggrcvpnlkgyftptvntvtdmdfiaequesfgpmliskfngsidslmq 832
QY 408 TANDTTYGLASVIYTKSLNRLGVSSALETGGVSINFPFIPETQTPFGMKQSGSGREL 467
DB 833 ranrteyglasgvftkdigkalnfadrieadgtvfvnynktdvaapfggkqgygkdlg 892

QY 468 EEGKAYLEPKTINI 482
: | | | :
Db 893 qealneyltkkcvtv 907

RESULT 15
AAG36240
ID AAG36240 standard; Protein: 424 AA.
XX
AC AAG36240;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 44384.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
KW
XX
OS Arabidopsis thaliana.
XX
EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-012180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130049.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.

```
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161892.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 27.5%; Score 700; DB 21; Length 424;
Best Local Similarity 39.3%; Pred. No. 8.3e-59;
Matches 169; Conservative 76; Mismatches 149; Indels 36; Gaps 9;

QY 70 TGAORAAACMLFADLAENAEKARLESPTGRPVSNMITHFDIPNMVSVFRYYAGWADKI 129
Db 2 tgeraklinkfadlienleelakdavgdklfqkyadlpataghrfrynagaadki 61

QY 130 AKCTFPDNGKPNWRY---EPMGYCAGIASWNAATFLYVGWKKIAPALAAACSFIFKASEKS 186
Db 62 hgetl-kmtrqslfgytlkepgvvgnlipwnfpsimfatkvapamaagctmrvkpaedt 120
```

```
QY 187 PLGVGLAPLFAEAGFPFGVQVFLTG-ARVTGEALASHMDIAKISFTRESVGGGAYKQAT 245
Db 121 sisalfyahlskeaglpdgvlnivtfgstagaaalashmdvdkvstfgstdvgrkinqaa 180
QY 246 LKSNMRRVTLELGEK-PTIVFNEAPLEROSGESAKD-----FSKFGQIWWPPSCLLVQW 298
Db 181 aasnkkvslelggkspllifndadidk-----aadlallgcfnkgeicvassrvfvqe 235
QY 299 G-----NLAEKFHCVRHGSGFCGQRMWLGONPLPKRTHGPFVDKSOYDRVLGNIDVGK 351
Db 236 gfydkvveklvek-----akdwlvqdpfdstarqgqvdkrfekilsyienhk 284
QY 352 -DTAQLLTGVGRKGDGKFAIEPTIFVNPCKSGKIWFEEIFGVPVLSIKTFKTEEAIAIAN 410
Db 285 negatlitggkaigdkgyfiqptifadvtedmklygdeifgpvmslmkftveegikcan 344
QY 411 DTTYGLASVIYTKSLNRGLRVSSALETGGVSINRPFIPETOTPFGGMKQSGSRELGEEG 470
Db 345 ntgyglaagilsqddidlintvarsikagiiwvncyfgfldcpvggykmsgncresgmda 404
QY 471 LKAYLEPKTI 480
Db 405 ldhylqtksv 414

Search completed: April 27, 2002, 07:45:33
Job time: 8114 sec
```

.....

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic ~ nucleic search, using sw model

Run on: April 26, 2002, 20:50:00 ; Search time 5001.19 seconds
(without alignments)
3950.968 Million cell updates/sec

Title: US-09-882-694A-4

Perfect score: 1464

Sequence: 1 atggcttcttcgctgacga.....tcacgtcaacatagatga 1464

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------------|--------------------|
| 1 | 232 | 15.8 | 2190 | 11 AK012213 | AK012213 Mus muscu |
| 2 | 158.8 | 10.8 | 653 | 10 BG810006 | BG810006 mgct002xf |
| 3 | 145 | 9.9 | 844 | 10 BF253896 | BF253896 HVSMEF000 |
| 4 | 143.4 | 9.8 | 673 | 10 BF850371 | BF850371 imagec_1 |
| 5 | 143.4 | 9.8 | 1024 | 9 AL135209 | AL135209 DKF2p762D |
| 6 | 140.6 | 9.6 | 603 | 10 BE405431 | BE405431 WHE1216_C |
| 7 | 139 | 9.5 | 1069 | 12 CNS0776W | AL432318 T3 end of |
| 8 | 138.2 | 9.4 | 660 | 10 BJ955705 | BJ955705 HVSME002 |
| 9 | 137.4 | 9.4 | 1100 | 10 BM461228 | BM461228 AGENCOURT |
| 10 | 131.8 | 9.0 | 1099 | 12 CNS0605B | AL410229 T7 end of |
| 11 | 131.2 | 9.0 | 656 | 9 BB626306 | BB626306 BB626306 |
| 12 | 130 | 8.9 | 395 | 10 BB428385 | BB428385 MTD006.D1 |
| 13 | 128.2 | 8.8 | 587 | 10 BE360262 | BE360262 DGL 62.B0 |
| 14 | 128 | 8.7 | 727 | 10 BF260792 | BF260792 HVSMEF002 |
| 15 | 127.6 | 8.7 | 612 | 9 AW565878 | AW565878 LGL_352.H |
| 16 | 126 | 8.6 | 614 | 9 AV917894 | AV917894 |
| 17 | 124.6 | 8.5 | 674 | 9 AW243051 | AW243051 xn28f12.x |

| | | | | | |
|----|-------|-----|------|-------------|--------------------|
| 18 | 124.2 | 8.5 | 508 | 9 A1158035 | A1158035 u326b09.r |
| 19 | 124 | 8.5 | 636 | 9 AV929175 | AV929175 AV929175 |
| 20 | 120.8 | 8.3 | 937 | 9 AL576537 | AL576537 AL576537 |
| 21 | 120.8 | 8.3 | 993 | 9 AL564722 | AL564722 AL564722 |
| 22 | 120.8 | 8.3 | 1047 | 12 CNS06NMY | AL407336 T7 end of |
| 23 | 120 | 8.2 | 656 | 10 BG356773 | BG356773 OV2_9.G07 |
| 24 | 120 | 8.2 | 1056 | 9 AL545580 | AL545580 AL545580 |
| 25 | 119.8 | 8.2 | 736 | 9 AV721915 | AV721915 AV721915 |
| 26 | 119.8 | 8.2 | 915 | 10 BF517895 | BF517895 603041971 |
| 27 | 119.2 | 8.1 | 1014 | 9 AL532910 | AL532910 AL532910 |
| 28 | 119 | 8.1 | 794 | 9 AL538753 | AL538753 AL538753 |
| 29 | 118.6 | 8.1 | 795 | 10 BF261094 | BF261094 HVSMEF002 |
| 30 | 118.2 | 8.1 | 794 | 10 BF740760 | BG740760 602632732 |
| 31 | 118.2 | 8.1 | 919 | 9 A1068478 | A1068478 mgae0002c |
| 32 | 118 | 8.1 | 643 | 10 BE362147 | BE362147 DGL 84.D0 |
| 33 | 117 | 8.0 | 703 | 10 BE362145 | BE362145 DGL 84.D0 |
| 34 | 117 | 8.0 | 723 | 9 AV721790 | AV721790 AV721790 |
| 35 | 116.6 | 8.0 | 755 | 10 BG967081 | BG967081 602833926 |
| 36 | 116.2 | 7.9 | 496 | 9 AA821260 | AA821260 vs15a12.r |
| 37 | 116 | 7.9 | 488 | 10 BG357200 | BG357200 OV2_11.A0 |
| 38 | 116 | 7.9 | 740 | 10 BG644873 | BG644873 EST506492 |
| 39 | 115.6 | 7.9 | 507 | 6 AW565284 | AW565284 LGL_332.G |
| 40 | 115.6 | 7.9 | 507 | 10 BM317596 | BM317596 LGL_214.G |
| 41 | 115.6 | 7.9 | 597 | 9 A1114304 | A1114304 GH1111.3 |
| 42 | 115.6 | 7.9 | 630 | 9 A1455414 | A1455414 LD33985.3 |
| 43 | 115.6 | 7.9 | 972 | 9 A1663335 | A1663335 u198a08.Y |
| 44 | 115 | 7.9 | 664 | 10 BG858411 | BG858411 1024057E0 |
| 45 | 115 | 7.9 | 739 | 10 BG915019 | BG915019 602813895 |

ALIGNMENTS

RESULT 1
AK012213

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AK012213 2190 bp mRNA linear HTC 19-JAN-2002
Mus musculus 11 days embryo whole body cDNA, RIKEN full-length
enriched library, clone:2700007F14:homolog to ALDH5YE
DEHYDROGENASE, MITOCHONDRIAL X PRECURSOR (EC 1.2.1.3) (ALDH CLASS
2), full insert sequence.

AK012213.1 GI:12848818

HTC; CAP trapper.

Mus musculus (strain:C57BL/6J) 11 days embryo cDNA to mRNA,
clone.lib:RIKEN full-length enriched mouse cDNA library
clone:2700007F14.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus; Mus.

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

99279253

10349636

2 (sites)

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

20499374

11042159

3 (sites)

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,

Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,

Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system-384-format

sequencing pipeline with 384 multicapillary sequencer

Db 1152 AGCAGTTTGAACGAATCCTGGCTACATCCGGCTGGGACAGAGGAGGCGCAAAAGCTTC 1211
 Qy 1070 tcactgctgttgtagaaggcgacaaaggattcggtattgaaccgaagatattgtca 1129
 Db 1212 TCTGTGGCGGGAGCGTTTGGGGAGCGCGCTTCTTCATAAACCCACAGTCTTCGGGG 1271
 Qy 1130 atcccacaacagcgacgacaaatttggttgagagatcttggcccgcttggccatta 1189
 Db 1272 ACCTTCAGGATGGCATGAGATCGCCAGGAGGAGATCTTGGCCCGTGACAGCTCTGT 1331
 Qy 1190 agacgttcaagacgggaagagagccattgagattgccaatgacacgactattgggctag 1249
 Db 1332 TCAAGTTCAAGAAGATCGAGGAAGTAATCCAGAGAGCCAAACACAGGATATGGCGTG 1391
 Qy 1250 cctcggtcattatacaaatctctcaacagggtctcggtctctgtogcgctcgaga 1309
 Db 1392 CTGGCGGTGTGTTACCCCGAGACCTGGACNAGGCCATCTACTCAGCGAGCCCTGCGAAG 1451
 Qy 1310 ccggtggtctcgatcaacttccctttatcccccagacacaaactcgtttggcgga 1369
 Db 1452 CTGGACGGGTGTGGTGAACACCTATAAATTTGTCACCTGCCACAGCCATTCGGAGGCT 1511
 Qy 1370 tgaacaatcggtcgaagcagagactagggcgaagagggctcaaggcgtacttggagc 1429
 Db 1512 TTAAGGAATCTGGCAATGGCAGGAGCTGGGGAGGACGGGCTCAGAGCCTACCGGAGG 1571
 Qy 1430 ccaagaccattaatccacgt 1451
 Db 1572 TGAAGACTGTCAACATCAAGGT 1593

RESULT 2

BG810006 653 bp mRNA linear EST 22-MAY-2001
 LOCUS mgct002xf07f Magnaporthe grisea Appressorium Stage cDNA Magnaporthe
 DEFINITION grisea cDNA clone mgct002xf07f 5', mRNA sequence.
 ACCESSION BG810006
 VERSION BG810006.1 GI:14180986
 KEYWORDS EST.
 SOURCE Magnaporthe grisea.
 ORGANISM Magnaporthe grisea.
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.
 1 (bases 1 to 653)
 AUTHORS Choi, W. and Dean, R. A.
 TITLE Construction and sequence analysis of an appressorium stage cDNA
 JOURNAL library in the rice blast fungus, Magnaporthe grisea
 COMMENT Unpublished (2001)
 Contact: Ralph A. Dean
 Fungal Genomics Laboratory
 North Carolina State University
 Campus Box 7251, Raleigh, NC 27695, USA
 Tel: 919-513-0020
 Fax: 919-513-0024
 Email: ralph_dean@ncsu.edu
 Seq primer: T3 primer (AATTAACCTCCTCAATAAGG).
 Location/Qualifiers
 1. .653
 /organism="Magnaporthe grisea"
 /strain="70-15"
 /db_xref="taxon:148305"
 /clone="mgct002xf07f"
 /dev_stage="Germinated conidia on appressorium-inductive surface"
 /note="Vector: pBlueScript SK(+) Vector; Site 1: EcoRI;
 Site 2: XhoI; The appressorium formation-specific cDNA
 library was constructed from conidia germinated for 5-8 hr
 on an inductive surface. The library contains over 55,000
 clones with average insert size of 1.5 kbp."

132 a 205 c 174 g 142 t
 BASE COUNT
 ORIGIN

FEATURES

source
 BP253896 844 bp mRNA linear EST 22-OCT-2001
 LOCUS HVSMEF0002G19f Hordeum vulgare seedling root EST library HVCNDA0007
 DEFINITION (Etolated and unstressed) Hordeum vulgare cDNA clone
 HVSMEF0002G19f, mRNA sequence.
 ACCESSION BP253896
 VERSION BP253896.2 GI:13116793
 KEYWORDS EST.
 SOURCE Barley.
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 1 (bases 1 to 844)
 AUTHORS Wang, R., Close, T. J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu
 , Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D. W., Fenton
 , R. B., Oates, R. and Main, D.
 TITLE Development of a genetically and physically anchored EST resource
 for barley genomics: Morex unstressed seedling root cDNA library
 JOURNAL Unpublished (2001)
 COMMENT On Nov 16, 2000 this sequence version replaced gi:11183001.
 Contact: Wing RA

Query Match 10.8%; Score 158.8; DB 10; Length 653;
 Best local Similarity 57.8%; Pred. No. 1.5e-31;
 Matches 365; Conservative 0; Mismatches 257; Indels 12; Gaps 4;

Qy 188 agggcccatgaagaagtctcacagggtgcacaaagcgggcggtgcattgaatttcgagg 247
 Db 6 AGGGCGAGTGGAGCAAAAGCTCTCTGGACAGCGAAGCAAGCTTCACCAACCTCGCG 65
 Qy 248 acctgcgcgaagaacacgcgaagaagctcgtctgagagtcgctgcacacgcggtagac 307
 Db 66 AACTTGTGAGAGAAACCTTTGACCTGCTGCGCGGTGAGTCTCTCGACAACGGAAGT 125
 Qy 308 cgggtcgtgatcactcatttcgacattccaaatggtctcgtggttcttcgtactatg 367
 Db 126 CCTTGGATGGCCAAAGGTGACGTGCGTGGCGCGGTGCC---TTAGGTACTAG 182
 Qy 368 caggctggcgcaagaatgcgcgggaagaccttcccgaggacacacgcggaag---ccga 424
 Db 183 GTGGTTGGCGCGCAAGATCGAGGCAAGACCATTTGACATTTGCCCGGACATGTTTCACT 242
 Qy 425 attgcattacgacgagatggggtgtgtctggtattgccagctggaacgcgacttttc 484
 Db 243 ACACAGAGTGTGACCCCTATCGGTGCTCGGTCAAAATATCCCTCGAAGTTCCTCTTC 302
 Qy 485 ttacgtcggctggaagatagcccccgccctcgccggtcgtctcttcacttcactcaag 544
 Db 303 TCATGCTGGCATGGAAGCTCGGACCTGCCCTGGCCACGGGTAAACCATCGTCTGAAGA 362
 Qy 545 cctcggaataatccccgctggggttctgggctcgtctctctctctgcagacgcggtat 604
 Db 363 CCGCGGACCAACACCCCTCTCAGCCCTTGTCTTGGCCAAACCTTGATCAAGGAGCGGTT 422
 Qy 605 tccctctcgtgagtgctgctcactcactgagacaga---gtgacgggtggaagcattgg 661
 Db 423 TCCCGCAGAGGTGCTGTCACATCATCTCTGGCTGGCAAGGTTCGGGTGCTGCCATCT 482
 Qy 662 cgtgcacatgacattgcgaagatcagcttcacagatctcgtggggtgacgcgcgag 721
 Db 483 CCGCACATGAGATCGACAAAGGTTCCTTTCATGCTCCACCGCTGTTGGCGCGGACCA 542
 Qy 722 tcaagcaagcaactcaagtcacacatgaagcgcgtcactctagaactgggggaaaa-- 779
 Db 543 TCATGAAGGCTGCGCCCTCGTCCAAACCTCAAGAGGTTCACCTTGAGCTTGGTGGCAAGT 602
 Qy 780 -gcaacacatcgtcttcaacgaagctcctctcga 812
 Db 603 CGCCTAACATTTATCTTCAACGACGCCGATATCGA 636

Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Total hg bases = 553
 Seq primer: AATTAACCTCCTCAAAAGG
 High quality sequence stop: 628.

FEATURES

Location/Qualifiers
 1. 844

/organism="Hordeum vulgare"
 /cultivar="Morex"
 /db_xref="taxon:4513"
 /clone="HVSMEF002G19f"
 /clone_lib="Hordeum vulgare seedling root EST library
 HVCN00007 (Etiolated and unstressed)"
 /tissue_type="Seedling root"
 /lab_host="TJ0121"

/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
 Seeds were surface sterilized then germinated under axenic
 conditions in the dark at room temperature on filter paper
 with water, nystatin and cefotaxime in covered
 crystallization dishes. Five-day old seedling roots were
 then harvested, total RNA was prepared, poly(A) RNA was
 purified, one primary unamplified cDNA library was made,
 and 1 million pfu were in vivo excised to give pBluescript
 SK(-) cDNA phagemids. These steps were performed in the TJ
 Close laboratory at the University of California,
 Riverside (Choi, Close, Fenton). Phagemids were plated and
 picked at the Clemson University Genomics Institute (CUGI)
 (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
 preparations, DNA sequencing and sequence analysis were
 performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates
 , Rambo, Main). The sequence has been trimmed to remove
 vector sequence and contains a minimum of 100 bases of
 phred value 20 or above. For more details on library
 preparation and sequence analysis see
 http://www.genome.clemson.edu/projects/barley. To order
 this clone see http://www.genome.clemson.edu/orders Also
 see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
 Genetically and physically anchored EST resources for
 barley genomics. Barley Genetics Newsletter 31:29-30.
 (http://wheat.pw.usda.gov/gppages/bgn/31/cover.html)."

BASE COUNT

193 a 229 c 259 g 160 t 3 others

Query Match 9.9%; Score 145; DB 10; Length 844;
 Best Local Similarity 52.6%; Pred. No. 8.4e-28;
 Matches 387; Conservative 0; Mismatches 338; Indels 11; Gaps 3;

Qy 645 gacgggtgaagcattggcgtgcacatggacattgacgaatcagcttcacaaatctgt 704
 Db 17 GCGCGCGCGCGCATCGCTCCACATGGACGTTGACAGCGTTGCTTCACTGGTCTGG 76
 Qy 705 cggcgtgacggcgcgcgaagcaaacactaaagtcacacatgaagcgcgtcaact 764
 Db 77 TGAAGTAGCCGCCCTCATATGAGGACATCGCCCGGAGCAACCTCAAGAGCGGTATCGCT 136
 Qy 765 aqaactgggggaaagcaaccatcgcttctcaacgaagctctctcgaacggcagtcggg 824
 Db 137 GGAGCTCGCGGCAAGTCGCCCTCTATAATCTTCATGACGCCGACGTCGACATGGCGGT 196
 Qy 825 ggaatcggaaggtttctcaaaaattcgggcaaaattgggtcccccctctctgtttgt 884
 Db 197 CGAGCTCTCAAGGCTTGGCCATATCTTCAACAAGGGAGAGGTTCG-----GTTGCGG 249
 Qy 885 agtgcattggggaaatttagcggagaaaattccatggagtcgcgtcattggtcattggagg 944
 Db 250 GGTCTCGTGTATCTTCAGGAAGGGATATACGACGAGTTCGTGAAGAAGGCTGTGGTGG 309
 Qy 945 ctgct-cagagatggttggccagaagaccattggaacccaagagcgcatggtcccttcg 1003

Db 310 CTGCCACAACCTGGAAAGTCGGAGATCGGTTCGATGTGGCCACCAACATGGGTCCCCAGG 369
 Qy 1004 tcgcaagtccagtagacagagagtcctgggtaacattgacgttgccaag---gataccg 1060
 Db 370 TCGACAAGGAGCAATTTGAGAGGGTCTTAAGGTACTATTGAGCAGCGCAAGACCGAGGAG 429
 Qy 1061 cgcagctcctcactggcgttggtagaaaaggcgcaaaagggattccgctcgaattgaacccagca 1120
 Db 430 CGACTCTCTCACCGCGCAAAACCTGCGCGGACAAAGGATACTACATTGAGCCTACCA 489
 Qy 1121 tatttgtaattcccaaacacagcgcaaaaatttggttgagagagatcttggcccgctct 1180
 Db 490 TATTTGACAGAGTCACGAAGGAGCATGAAGTCGCTCAAGACGAGATCTTCGCCCCCGTGA 549
 Qy 1181 tgtccattaaagcgttcaagacgggaagagggccattgagattgccaatgacacgacct 1240
 Db 550 TGTCCCTCATGAAGTTCAAGACGGTCTGATGAGGCGATAGAGAAGGCCCAACTGCACCAAGT 609
 Qy 1241 atgggctagcctcggctcatttataccaaatctctcaacaggggtctcsgtctcgtcgtg 1300
 Db 610 ACGGCTCGCGCGNCGGGATCATACCAAGAACTTGGACATCGCCCAACAAAGGTGTCAAGT 669
 Qy 1301 cgtctgagacccggtggcgtctcgcatacaacttccccctttatccccgagacacaaactccgt 1360
 Db 670 CGGTTCGCGCGGACACAGTGTGTTGAAGTCTACTTCCGCTCGACCCCGAGGCGCCTT 729
 Qy 1361 ttgcccggcatgaaaca 1376
 Db 730 CGGCGGTACAAAGAGA 745

RESULT 4

BI850371 BI850371 673 bp mRNA linear EST 10-OCT-2001
 LOCUS IMAGEC_12_2000/smd97bdrb1.y1 NIH_MGC_84 Homo sapiens cDNA clone
 DEFINITION IMAGE:4360588 5', mRNA sequence.

ACCESSION BI850371
 VERSION BI850371.1 GI:16003858
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Carnivora; Homnidae; Homo.
 Kalle, P.I., Harsch, T.J., Polta, P.A., Nelson, D.O., Sanders, C.G. and
 Prange, C.K.

TITLE

The I.M.A.G.E. Consortium quality control effort: clone
 resequencing for verification

JOURNAL

Unpublished (2001)

COMMENT

Other ESTs: BF989915

Contact: Prange CK

The I.M.A.G.E. Consortium

Lawrence Livermore National Laboratory

Livermore, CA, USA

Email: help@image.llnl.gov

This read has been verified (found to hit its original self in the
 correct orientation), as part of the I.M.A.G.E. Consortium quality
 control effort. High quality sequence is defined as having 100 or
 more base pairs with a phred quality value of 20 or greater, where
 a sliding window of 4 base pairs marks the beginning and end of the
 sequence. For information on obtaining this clone, please contact
 info@image.llnl.gov.

Plate: LLAM10002 row: a column: 5

Seq primer: ml3rp1

High quality sequence stop: 673.

Location/Qualifiers

1. 673

FEATURES

source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4360588"
 /clone_lib="NIH_MGC_84"
 /tissue_type="adrenal cortex carcinoma, cell line"

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.

1 (bases 1 to 603)

Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han
P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J.,
Seaton, C.D. and Tong, J.C.
The structure and function of the expressed portion of the wheat
genomes

Unpublished (2000)

Contact: Olin Anderson

US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818

Email: andersn@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20

Seq primer: Strategene SK primer.

FEATURES
source

Location/Qualifiers
1..603
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone_lib="WHE1216_C09_F18"
/tissue_type="Root"
/dev_stage="Five day old etiolated seedling root cDNA library"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site_1: EcoRI; Site_2: XhoI; Seeds were surface-sterilized
, germinated and grown aseptically in the dark at room
temperature on filter paper with water, nystatin and
cefotaxime in covered crystallization dishes. Roots were
harvested. The tissue, total RNA, and poly(A) RNA were
prepared. A cDNA library was made, and the cDNA clones
were in vivo excised to give phagescript phagemids in the
TJ Close lab (Choi, Close, Fenton) at the University of
California, Riverside. Plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."

BASE COUNT 147 a 142 c 187 g 127 t
ORIGIN

Query Match 9.6%; Score 140.6; DB 10; Length 603;

Best Local Similarity 56.2%; Pred. No. 1e-26;

Matches 285; Conservative 0; Mismatches 219; Indels 3; Gaps 1;

QY 949 cagagatggcttgccagagaccattggaacccaagagagcgatgctcccttgcgac 1008

DB 17 CAGAACTGGNAAGTCGGAGACCCGTTGCGAGTCGCCACCAACATGGTGCCCGAGTTGAT 76

QY 1009 aagtcacagtacacagagcttggtggaacattgacgttggaag---gataccgcgacg 1065

DB 77 AAGGAGCAATTTCAAGGGTCTTAAGCTATGACGTGCGAAGACGAGGGGGCGACT 136

QY 1066 ctctcactggccttggttagaagggcgacaagggatcgattgaacgcagcatattt 1125

DB 137 CTGCTGACCGGGCGCAACGTCGGCGGCAAAAGGATACATGAGCCTACCATATTT 196

QY 1126 gtcaatcccaaacagcgacgaaaaatttggttgagagatatgtggcccgcttctgac 1185

DB 197 GCAGATGTCAGGAGACATGAAGATCGCTCAAGATGAGATCTTTGGCCCTGTGATGTC 256

QY 1186 attagacgttcaagcgaagaagagggcattgagattgccaatgacacgaacttatggg 1245

DB 257 CTCATGAAGTTCAAGACGCTCGATGATGCGGATAGACAGGCCCACTGCACGAAGTACGGG 316

QY 1246 cttagctcggtcatttataccaaatctctcaacaggggtctcgtgtctcgtcgcgctc 1305

DB 317 CTGGCGCGCGGGATAATCAACCAAGAACTTTGGACATCGCCCAACCGGCTGTCGAGGTGCGGTT 376

QY 1306 gagaccgggtggcgtctcgatcaacttccctttatccccggagacacaaactcgtttggc 1365

DB 377 CGAGCGGGTACCGTGTGGGTGAATTTGTTACTTCGCTTCGATCCCGAGCGCCCTTTGGC 436

QY 1366 ggcataaacaatcggtctcagcgagagctagctcgaagaagggctcaagcgctacttg 1425

DB 437 GGGTACAAGATGACGGCGCTTCGCCGGGACGAGGGGATGATGGCATCGACAAGTACATG 496

QY 1426 gagcccaagaccattaatatcccgctc 1452

DB 497 CATGTCAAGAGCGCTCATCACCGGAGTC 523

RESULT 7

CNS0776W 1069 bp DNA linear GSS 08-JUL-2001
T3 end of clone BB0AA009F08 of library BB0AA from strain CBS 4732
of Pichia angusta, genomic survey sequence.

ACCESSION AL432318
VERSION AL432318.1 GI:12215732

KEYWORDS GSS.
SOURCE Pichia angusta.
ORGANISM Pichia angusta

REFERENCE Eukaryota: Fungi; Ascomycota: Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Pichia.

AUTHORS 1 (bases 1 to 1069)
Blandin, G., Llorente, B., Malpertuy, A., Wincker, P., Artiguenave, F.,
and Dujon, B.

TITLE Genomic exploration of the hemiascomycetous yeasts: 13. Pichia
angusta

JOURNAL FEBS Lett. 487 (1), 76-81 (2000)

MEDLINE 20584723

REFERENCE 2 (bases 1 to 1069)
Soucie, J.L., Aglie, M., Artiguenave, F., Blandin, G.,
Boletin-Fukuhara, M., Bon, E., Bottier, P., Casaregola, S.,
de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,
Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,
Saurin, W., Tekala, F., Toffano-Nloche, C., Wesolowski-Louvel, M.,
Wincker, P., and Weissenbach, J.

TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies

JOURNAL FEBS Lett. 487 (1), 3-12 (2000)

MEDLINE 20584711

REFERENCE 3 (bases 1 to 1069)
Genoscope.

AUTHORS Direct Submission

JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
source

1..1069
/organism="Pichia angusta"
/strain="CBS 4732"
/db_xref="taxon:4905"
/clone="BB0AA009F08"
/clone_lib="BB0AA"
/note="end : T3"
<12..>1067

misc_feature
/note="similar to Saccharomyces cerevisiae ORF YER073w [
ALD5 : aldehyde dehydrogenase (NAD+), mitochondrial]"

misc_feature
/evidence=not_experimental
<15..>1067
/note="similar to Saccharomyces cerevisiae ORF YOR374w [
ALD5 : aldehyde dehydrogenase (NAD+), mitochondrial]"

| | | | |
|------------|------|---|------|
| QY | 1127 | tcaatcccaaacaggagcagaataattggtttggaggagatatcttggccccgccttgtcca | 1186 |
| Db | 901 | CCGAGCTGTACCCAGCATGAGCATAACAAGGAGGAGATCTTTGGACCATTTCGCCGTGA | 960 |
| QY | 1187 | ttaagacgttcgaagacaagaagacagagccacttgagattgccaatgacacgaccttatggc | 1246 |
| Db | 961 | TCAGCCCATCAAGACCGCCGACGAGGATCGAGTTGGCTAACGACTCCGAATACGGAC | 1020 |
| QY | 1247 | tagcctggg | 1255 |
| Db | 1021 | TTCGTGCAG | 1029 |
| RESULT | 8 | | |
| BI955705 | | | |
| LOCUS | | | |
| DEFINITION | | | |
| ACCESSION | | | |
| VERSION | | | |
| KEYWORDS | | | |
| SOURCE | | | |
| ORGANISM | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| COMMENT | | | |
| FEATURES | | | |
| source | | | |

ORGANISM

Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE

1 (bases 1 to 612)

AUTHORS

Cordonnier-Pratt M.-M., Gingle, A., Marsala, C. and Pratt, L.H.

TITLE

An EST database from Sorghum: light-grown seedlings

JOURNAL

Unpublished (2000)

COMMENT

Contact: Cordonnier-Pratt MM

Department of Botany

The University of Georgia

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 542 1805

Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.

Seq primer: T7

High quality sequence start: 27

High quality sequence stop: 612

POLYA=Yes.

FEATURES

Location/Qualifiers

1..612

/organism="Sorghum bicolor"

/db_xref="taxon:4558"

/clone_lib="Light Grown 1 (LGI)"

/note="Organ: 10- to 14-day-old light-grown (greenhouse)

seedlings; Vector: Lambda Zap; Site1: XhoI; Site2: EcoRI

; The library was made from poly-A RNA in the cloning

vector lambda ZAP II. Clones to be sequenced were

prepared by mass excision."

154 a 163 c 151 g 144 t

BASE COUNT

ORIGIN

Query Match 8.7%; Score 127.6; DB 9; Length 612;
Best Local Similarity 55.3%; Pred. No. 3.1e-23;
Matches 269; Conservative 0; Mismatches 214; Indels 3; Gaps 1;

Qy 970 ccattggaccacgaaggacgcatgtcccttcgtcgacaaagtcacagacagagtc 1029

Db 1 CCATTGCCCGCAGGAACCTTCCAGGGTCTCAGGTTCCAGCTTCAGTTCGACCGCATC 60

Qy 1030 ttgggttaacattgacttggaagga---tacgcgcagctcctcactgcggttgtaga 1086

Db 61 ATGAGCTACATTGAGGATGTAGAAGTCTGGTGCCACTGTGCAAAACCGGTGGTAAGCGC 120

Qy 1087 aaggcgacaaaggattcgcatgaaccgcagcatattgtcaatcccaaacccaggcagc 1146

Db 121 AAGGGCGACAAGGGTTACTTCATCGAGCGCCACCATCTTCTCCAACGTCACCGAGGACATG 180

Qy 1147 aaaaattgtttagagatctttggcccgctctgtccattaaagcagttcaagacggaa 1206

Db 181 AAGATTGAGCAAGAGAGATCTTGGCCCGCTTGCACAACTCCCAAGTTCAAGAACAG 240

Qy 1207 gaagagccattgagattgccaatgacacgacttatgggttagcctcggtcatttatacc 1266

Db 241 GAGGAGTTCATCAAGATTGGCAACTCGACCTCATACGGTCTTGTCTGCCGTCCACACC 300

Qy 1267 aaattctcaagagggttcctcgctctgtcggtcgctcgagacgggtgggtctcgatc 1326

Db 301 ACCAACCTCACCCGCCCTCGAGGTTGCCAATGCGCTCAGGGCCGGAACCTCTCGGGTC 360

Qy 1327 aacttccctttatccccgagacacaaactccgtttggcggtgatgaacaatcgggtca 1386

Db 361 AACACATACACGCTCTTCACTGGTCACTGCCATTCCGTTGGTTACAAGGAGTCTGGTATT 420

Qy 1387 ggcagagagctagcggaagagggtcctcaaggcgctacttgagcccaagaccattaatc 1446

Db 421 GGCGGTGAGTGGGTGAGCGCGCTGGAGAACTACGTCCTCAGACCAAGACGGTCTCGATC 480

Qy 1447 cactgc 1452

Db 481 CGTCTC 486

Search completed: April 26, 2002, 20:50:14
Job time: 5278 sec

```
Db 455 ctatgtcatctctactcctggtggtgatttggaacatggtctcctcaaatgtctccggtattatgc 514
Qy 369 aggtcgggcccagacatgcgcggaaaagacctttccc---gaggacaacggcaagccgaa 425
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 515 cggctgggctgataagtagaccacggaaaaccatcccatctgacggagactttctcagcta 574
Qy 426 ttggcgttacgagcagatgggggtgtgtgctggtattgccagctggaacggagactttct 485
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 575 cacacgcatgaacctgtgggggtgtgctgggcagatcctccgtggaatttcccgcctct 634
Qy 486 ttacgtcgctggaagatagcccgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 545
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 635 gatgaagcatggaagctggcccgacgttggaactggaaacgtggttgatgaaggt 694
Qy 546 ctgcgagaaatcccgcctg999gcttctg999cctcctcctcctcctcctcctcctcctc 605
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 695 agctgagcagatcacccctcacgcctctatgtgccaacctgatcaaggagcgtgctt 754
Qy 606 cctcctgagctgctgacgttccctcactggagcagagtagcgggtgaa--gcattggc 662
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 755 tcccctggctggccaaacatctgtcctggatttggccccacggctgg999cgcgcattgc 814
Qy 663 gtccacatggacattgcgaagatcagcttcacaagatctgtcgcggtggccgcgcgcgt 722
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 815 ctcccatgaggtatggacaagtggtcattccacaggtccactgagattggccgcgtaat 874
Qy 723 caagcaagaacactcaagtcacaacatgaagcgcctcactctagaaactgggggaaagcc 782
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 875 ccaggtgtcgtg99gagcagcaactcaagagatgacctggagctggggggggaadag 934
Qy 783 aacctgtcttcaacagagctcctctcgaaacggcagtcgggggaatcggcgaagattt 842
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 935 cccaacatcatcgtcagatgcgatggtggtggtggtggtggtggtggtggtggtggtggt 994
Qy 843 ctcaaaattcgggcaaatgttggttccccctcctcctctgttctgtagtgcgaatggggaaatt 902
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 995 cctgtttctcaacca-----gggcagtgctgtgctgctgcgcgcgcgcgcgcgcgcgc 1048
Qy 903 agcggagaaattccatggagtcgcgtcatgttcatttgaggctgtcagagatgctctgg 962
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1049 ggaggacatctatgattggttggagcggagctttgcccgcccaagctcgggtggt 1108
Qy 963 ccagaaacctatggaaacccaagagacgacatggttcccttcctgcacaaagtcaccaagca 1022
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1109 cgggaacctcttgatagaacagcagcagggcgccgaggtggatgaaactcagtttaa 1168
Qy 1023 cagagcttgggttaacattgactgtggcaag---gataccgcgcgcgcgcgcgcgcgcgc 1079
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1169 gaagatcctcgcgtacatacaacacggggaagcaagagggggcgaaagctgtgtgggtg 1228
Qy 1080 tggtagaaaggcgacaagggatttcgcgattgaacgcgaagatatttgcatacccaaac 1139
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1229 gggcattgctgtgacccgtgttacttcatccagcccactgtgttggagatgtgcagga 1288
Qy 1140 aggcagcaaaatttggttggagagatcttggccgcctctgtccattgaagacttcaa 1199
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1289 tggcatgaccatgcgcaaggagagatcttcggccagtgatgcagatcctgaaattcaa 1348
Qy 1200 gacgaaagagggccattgagattgccaatgacacgacttatggctagcctcgtggtcat 1259
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1349 gaccatagagggtgtgttggagagccaacaattccacgtacgggctggccgcagctgt 1408
Qy 1260 ttataccaactctcaacaggggtctcgtgtctcgtcgcgcgcgcgcgcgcgcgcgcgc 1319
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1409 ctcaacaaggatttggacaaggccaattacctgtcccaaggccctcccgggcgcaactgt 1468
Qy 1320 ctcatcaacttcccctttatccccgagacacaaactccgttttggcgcatgaacaatc 1379
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1469 gtgggtcaactgtatgatgttggagcccaagctcacccttgggtggtcaagatgtc 1528
Qy 1380 gggctcaggcagagctagggcgaagagggctcaaggctacttggagcccaagaccat 1439
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1529 ggggagtgccggaggttgggcgagtacgggctgcaggcatatcactgaagtgaactgt 1588
```

```
Qy 1440 taatalccacgt 1451
    | | | | |
Db 1589 cacagtcaagt 1600
```

RESULT 2

```
US-10-106-698-2021/c
; Sequence 2021, Application US/10106698
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypep
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 2021
; LENGTH: 994
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-106-698-2021
```

```
Query Match 6.1%; Score 90; DB 6; Length 994;
Best Local Similarity 55.9%; Pred. No. 2.4e-17;
Matches 171; Conservative 0; Mismatches 135; Indels 0; Gaps 0;
```

```
Qy 1146 caaaatttggttgaggagatctttggcccgctgtgtccattaaagagcttcaagacgga 1205
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 853 CAAGATGCCCAAGGAGGAGATCTTCGGCCAGTCAGTGCAGATTCCTGAAAGTTCAAGACCAT 794
Qy 1206 aagaagggcattgagattgccaatgacagcacttatggctagctcgggtcatttatac 1265
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 793 AGAGGAGGTTGTGGGAGAGACCAACAATTCACGTACGGGCTGGCGGACGCTGCTTTCAC 734
Qy 1266 caaatctctcaacaggggtctcgtgtctcgtcgcgcgtcgcagacgggtggcgtctcgt 1325
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 733 AAAGGATTGGACAAGGCCAAATACCTGTCCAGGCCCTCCAGGGGGGACACTGTGTGGGT 674
Qy 1326 caacttcccttattccccgagacacaaactcctgttggcgcgcgcgcgcgcgcgcgcgc 1385
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 673 CAACCTGCTATGATGTGTTGGAGCCCACTACACCTTTGGTGGCTACAAGATGTCGGGGAG 614
Qy 1386 aggcagagcttagcgcgaagaggctcaaggctacttggggcccaagaccattaatat 1445
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 613 TGGCGGGAGTTGGCGGAGTAGCGGCTCGAGGCATACACTGAAGTGAAACACTGCACAGT 554
Qy 1446 ccacgt 1451
    | | | | |
Db 553 CAAAGT 548
```

RESULT 3

```
US-09-975-254-19603
; Sequence 19603, Application US/09975254
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Heck, Gregory R.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(15309)B
; CURRENT APPLICATION NUMBER: US/09/975,254
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US/09/263,191
; PRIOR FILING DATE: 1999-03-05
```



```
; NUMBER OF SEQ ID NOS: 31255
; SEQ ID NO 19603
; LENGTH: 245
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 700957115H1
US-09-975-254-19603

Query Match      2.9%; Score 42.8; DB 5; Length 245;
Best Local Similarity 55.1%; Pred. No. 0.0015;
Matches 103; Conservative 0; Mismatches 83; Indels 1; Gaps 1;

QY 396 gaccttcccgagacacgcaagcgaattggcgtttacgagccgagtggtgtgtac 455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 53 gctcttcccatgacacattcaagagttatgttcaaggagccgagtggtgtgtgc 112
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 456 tgggtattccagctggaacgcgacttttcttacctgctggaagatagcccccgcct 515
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 113 tttaataactccttggaattatcncgtgtgatggtacg-tggaaagttgctcgtct 171
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 516 cgcgcgcgcgtcctcttcattcctcaaaagcctcgagaaatcccgcgtggtctggg 575
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172 ggcgcgcgcgtgctgctgaattgaagccctctgagtcgtggaatgtgacatgtttgga 231
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 576 cctcgct 582
    ||| |||
Db 232 gctcgct 238

RESULT 4
US-09-975-254-11784
; Sequence 11784, Application US/09975254
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Heck, Gregory R.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(15309)B
; CURRENT APPLICATION NUMBER: US/09/975,254
; PRIOR FILING DATE: 2001-10-12
; PRIOR FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 31255
; SEQ ID NO 11784
; LENGTH: 247
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 700905254H1
US-09-975-254-11784

Query Match      2.8%; Score 41; DB 5; Length 247;
Best Local Similarity 57.3%; Pred. No. 0.0052;
Matches 71; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 1078 gttggtagaaggcgacagggattcgcgattgaaccagcagatattgtcaatcccaaa 1137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110 ggtgttcaatgganganaatgacttatttgaaccgacaaatattgtggtgacgcga 169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1138 ccaggcagcaaatgtgttgagagatctttggcccgctgttcccaatgaagcgttc 1197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 170 cttgattcagcagtcagtgctgngaaatctttggccctgtacttccaaataacccctg 229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1198 aaga 1201
    |||
Db 230 gaaa 233

RESULT 5
US-09-975-254-24522
; Sequence 24522, Application US/09975254
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Heck, Gregory R.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(15309)B
; CURRENT APPLICATION NUMBER: US/09/975,254
; PRIOR FILING DATE: 2001-10-12
; PRIOR FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 31255
; SEQ ID NO 24522
; LENGTH: 285
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 700963635H1
US-09-975-254-24522

Query Match      2.7%; Score 39; DB 5; Length 285;
Best Local Similarity 52.7%; Pred. No. 0.022;
Matches 107; Conservative 0; Mismatches 95; Indels 1; Gaps 1;

QY 439 ccgatgggggtgtgtgctggtattgccagctggaaogagactttcttttaoagctgg 498
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 cccatcgaggtgtgttcatttaattactccatggaactatctctattgtatgctacatgg 119
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 499 aagatagccccccctcgccgcgctgctctcttcattcattcaaaagcctcgagaaatcc 558
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 120 aaagtgtcctctctctagctgctggttg-tacacaattatgaagccatcgaattggca 178
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 559 ccgctggggcttctggcctcgctctctctctcttcagaaagccgagttccctctgagctc 618
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 tctgtgacctgtttggagctggtgaaatgacagaaagttggcctctctccagggata 238
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 619 gtgcagttctcactgagcaag 641
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 ttaacaattgttactggtagg 261
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
US-09-975-254-22618
; Sequence 22618, Application US/09975254
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Heck, Gregory R.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(15309)B
; CURRENT APPLICATION NUMBER: US/09/975,254
; PRIOR FILING DATE: 2001-10-12
; PRIOR FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 31255
; SEQ ID NO 22618
; LENGTH: 254
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 700961292H1
US-09-975-254-22618

Query Match      2.6%; Score 37.6; DB 5; Length 254;
Best Local Similarity 54.3%; Pred. No. 0.054;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 1108 attgaaccgacgatattgtcaatcccaaccagcagcaaaatttggttgagagatc 1167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 86 atttgcccttctgctactggacaattgttagccagacatgagaaatgcatgggaagacca 145
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1168 ttggccccctcttctgtccattaaagacgttcaagacggaagagagccattgagattgcc 1227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


| | Matches | 71; | Conservative | 0; | Mismatches | 65; | Indels | 0; | Gaps | 0; |
|----|---------|---|--------------|----|------------|-----|--------|----|------|----|
| QY | 507 | ccccgcctcgccgcggctgctcttcattcctcaaaagccctcgagaaaatccccgctggg | 566 | | | | | | | |
| Db | 6065 | CCGCGGCCCCACCTATGTCTGACCTTAACTTGAAGACCAAGAGAGAGAGGACGGAAGTG | 6006 | | | | | | | |
| QY | 567 | cgtcttgggcctcgctctctctcttcgcagaagccggattccctcctggagtcgtgcagtt | 626 | | | | | | | |
| Db | 6005 | GATCCTGGCAGCCGTGGCGCTGCTCCTACAGGTTTAGCTCGCTCCTGCCTCAGAGCCAC | 5946 | | | | | | | |
| QY | 627 | cctcactggagacga | 642 | | | | | | | |
| Db | 5945 | ACTCCCTGGGCTGGA | 5930 | | | | | | | |

Search completed: April 27, 2002, 01:36:31
Job time: 17124 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 01:33:16 ; Search time 8172.36 Seconds
(without alignments)
3875.685 Million cell updates/sec

Title: US-09-882-694A-4
Perfect score: 1464
Sequence: 1 atgggtcttcgctgaacga.....tccacgtcaacatagtgta 1464

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 21979536 seqs, 10817449327 residues 43959072
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*

1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*

2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*

3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*

4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*

5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*

6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*

7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*

8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*

9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*

10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*

11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*

12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*

13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*

14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*

15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*

16: /cgn2_6/ptodata/2/pna/US092_COMB.seq:*

17: /cgn2_6/ptodata/2/pna/US093_COMB.seq:*

18: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*

19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*

20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*

21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*

22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*

23: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*

24: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*

25: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*

26: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*

27: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*

28: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*

29: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*

30: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*

31: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*

32: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*

33: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*

34: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*

35: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*

36: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*

37: /cgn2_6/ptodata/2/pna/US100_COMB.seq:*

38: /cgn2_6/ptodata/2/pna/US101_COMB.seq:*

39: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*

40: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*

41: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*

42: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*

43: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*

44: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*

45: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*

46: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*

47: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*

48: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*

49: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*

50: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*

51: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*

52: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*

53: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*

54: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*

55: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*

56: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*

57: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*

58: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*

59: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*

60: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*

61: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*

62: /cgn2_6/ptodata/2/pna/US6023_COMB.seq:*

63: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*

64: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*

65: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*

66: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*

67: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*

68: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*

69: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*

70: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*

71: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*

72: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*

73: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*

74: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*

75: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Query Length | IDB | Description |
|------------|-------|---------|--------------|-----|--|
| 1 | 1464 | 100.0 | 1464 | 17 | US-09-351-224-4
Sequence 4, Appli |
| 2 | 1464 | 100.0 | 1464 | 17 | US-09-351-224E-4
Sequence 4, Appli |
| 3 | 1464 | 100.0 | 1464 | 17 | US-09-351-823-4
Sequence 4, Appli |
| 4 | 1464 | 100.0 | 1464 | 26 | US-09-677-488-4
Sequence 4, Appli |
| 5 | 1464 | 100.0 | 1464 | 26 | US-09-677-488A-4
Sequence 4, Appli |
| 6 | 1464 | 100.0 | 1464 | 26 | US-09-677-682-4
Sequence 4, Appli |
| 7 | 1464 | 100.0 | 1464 | 26 | US-09-677-682A-4
Sequence 4, Appli |
| 8 | 1464 | 100.0 | 1464 | 26 | US-09-677-682B-4
Sequence 4, Appli |
| 9 | 1464 | 100.0 | 1464 | 33 | US-09-882-694A-4
Sequence 4, Appli |
| 10 | 1464 | 100.0 | 1464 | 33 | US-09-882-694A-4
Sequence 4, Appli |
| 11 | 267.6 | 18.3 | 1491 | 10 | US-08-612-853-1
Sequence 1, Appli |
| 12 | 250.6 | 17.1 | 1491 | 18 | US-09-404-520-27986
Sequence 27986, A |
| 13 | 245.8 | 16.8 | 2022 | 20 | US-09-533-559-5519
Sequence 5519, Ap |
| 14 | 245.4 | 16.8 | 1710 | 20 | US-09-533-559-14
Sequence 14, Appl |
| 15 | 241.6 | 16.5 | 9589 | 1 | PCT-US01-17253-89
Sequence 89, Appl |
| 16 | 240.4 | 16.4 | 1551 | 61 | US-60-226-176-419
Sequence 419, App |
| 17 | 240.4 | 16.4 | 1551 | 62 | US-60-233-468-419
Sequence 419, App |
| 18 | 240.4 | 16.4 | 1551 | 70 | US-60-313-371-419
Sequence 418, App |
| 19 | 240.4 | 16.4 | 2041 | 61 | US-60-226-176-418
Sequence 418, App |
| 20 | 240.4 | 16.4 | 2041 | 62 | US-60-233-468-418
Sequence 418, App |
| 21 | 240.4 | 16.4 | 2041 | 70 | US-60-313-371-418
Sequence 418, App |
| 22 | 240.4 | 16.4 | 2288 | 60 | US-60-213-360-6928
Sequence 6928, A |
| 23 | 240.4 | 16.4 | 2308 | 56 | US-60-172-373-11821
Sequence 11821, A |
| 24 | 240.4 | 16.4 | 2552 | 25 | US-09-644-867-7895
Sequence 7895, Ap |
| 25 | 240.4 | 16.4 | 2552 | 25 | US-09-644-871-9732
Sequence 9732, Ap |
| 26 | 240.4 | 16.4 | 2552 | 25 | US-09-652-126-9381
Sequence 9381, Ap |
| 27 | 240.4 | 16.4 | 2552 | 25 | US-09-652-128-10113
Sequence 10113, A |
| 28 | 240.4 | 16.4 | 2552 | 25 | US-09-652-911-10185
Sequence 10185, A |
| 29 | 240.4 | 16.4 | 2552 | 25 | US-09-652-914-9132
Sequence 9132, Ap |
| 30 | 240.4 | 16.4 | 2552 | 25 | US-09-652-916-9779
Sequence 9779, Ap |
| 31 | 240.4 | 16.4 | 2552 | 27 | US-09-698-010-12416
Sequence 12416, A |

; Sequence 4, Application US/09351224E
; GENERAL INFORMATION:
; APPLICANT: Duivick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE OF INVENTION: Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/351,224E
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: aldehyde dehydrogenase, fully spliced cDNA
US-09-351-224E-4

Query Match 100.0%; Score 1464; DB 17; Length 1464;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggtctcttcgctgagaaacacagagtgagctcttcacacaaatgaattgcttcc 60
Db 1 atggtctcttcgctgagaaacacagagtgagctcttcacacaaatgaattgcttcc 60

QY 61 tcaaggggtccgagagatgaacgtccagacccggtggagcaatccacggttgcact 120
Db 61 tcaaggggtccgagagatgaacgtccagacccggtggagcaatccacggttgcact 120

QY 121 gatgttcagctggccaaacgcccagatgacagatgacagtcgctgagcgttcgtaggcg 180
Db 121 gatgttcagctggccaaacgcccagatgacagatgacagtcgctgagcgttcgtaggcg 180

QY 181 gtcaaaaggccatgggaagatgacaggtgcacacgcccggcgctgcatgcttaag 240
Db 181 gtcaaaaggccatgggaagatgacaggtgcacacgcccggcgctgcatgcttaag 240

QY 241 ttccggagacctgcgagagaaacccgagagctgcgtctgagcgttcgtaggcgctccacc 300
Db 241 ttccggagacctgcgagagaaacccgagagctgcgtctgagcgttcgtaggcgctccacc 300

QY 301 ggtagaccggtgctgagatgacactatttcgacattccaaacatggtctccggttttcgc 360
Db 301 ggtagaccggtgctgagatgacactatttcgacattccaaacatggtctccggttttcgc 360

QY 361 taactatcgagctggccagagatgacagacacacacacacacacacacacacacacac 420
Db 361 taactatcgagctggccagagatgacagacacacacacacacacacacacacacacac 420

QY 421 ccgaattggcgttacgagccgagtggtgtgctgtgtatttcgagctggagacgcgact 480
Db 421 ccgaattggcgttacgagccgagtggtgtgctgtgtatttcgagctggagacgcgact 480

QY 481 ttcttttaagctggctggaagatagccccccgctgcgcgcgcgcgcgcgcgcgcgcgcgc 540
Db 481 ttcttttaagctggctggaagatagccccccgctgcgcgcgcgcgcgcgcgcgcgcgcgc 540

QY 541 aaagcctcggagaaatcccgctggcgcttctgggcctgcgtccctctcttcgcaagacc 600
Db 541 aaagcctcggagaaatcccgctggcgcttctgggcctgcgtccctctcttcgcaagacc 600

QY 601 ggaattccctctgagctgtagtctcacttgagacagcagtgacgggtgaagcattg 660
Db 601 ggaattccctctgagctgtagtctcacttgagacagcagtgacgggtgaagcattg 660

QY 661 gcgtcgacatgacattcgcgaagatcagcttcacaaagatctgtcgcgggtggcgcgcgc 720
Db 661 gcgtcgacatgacattcgcgaagatcagcttcacaaagatctgtcgcgggtggcgcgcgc 720

QY 721 gtcaagcaagcaacactcaagtccaacatgaagcgcgtcactctagaactgggggaaag 780
Db 721 gtcaagcaagcaacactcaagtccaacatgaagcgcgtcactctagaactgggggaaag 780

QY 781 caaacatcgtcttcaacgaagctctctcgaacgcgagtcgggggaatcggcaagat 840
Db 781 caaacatcgtcttcaacgaagctctctcgaacgcgagtcgggggaatcggcaagat 840

QY 841 ttctcaaaattcgggcaaaatttgggtcccccctctctgttctagtcaatggggaat 900
Db 841 ttctcaaaattcgggcaaaatttgggtcccccctctctgttctagtcaatggggaat 900

QY 901 ttacgcgagaaattccatgagtcgctgctcatttggagcgtgtcgaagatggcgtt 960
Db 901 ttacgcgagaaattccatgagtcgctgctcatttggagcgtgtcgaagatggcgtt 960

QY 961 ggccagaacccattggaaacccaagagacgcgtgtccctctcgtcgacaagtccacgtac 1020
Db 961 ggccagaacccattggaaacccaagagacgcgtgtccctctcgtcgacaagtccacgtac 1020

QY 1021 gacagagctctgggttaacattgacgttggcaagagataccgcgcagctcctcactggcgtt 1080
Db 1021 gacagagctctgggttaacattgacgttggcaagagataccgcgcagctcctcactggcgtt 1080

QY 1081 ggtagaaaggcgacaaaggattcgcgttgaacacgcagataattgtcaatcccaaacca 1140
Db 1081 ggtagaaaggcgacaaaggattcgcgttgaacacgcagataattgtcaatcccaaacca 1140

QY 1141 ggacgaaatttgggttggagagatctttggcccgtctgtccattaaagacgttcaag 1200
Db 1141 ggacgaaatttgggttggagagatctttggcccgtctgtccattaaagacgttcaag 1200

QY 1201 acggaagagagccattgagattgccaatgacacgacttatggcgtgacgtcgtcatt 1260
Db 1201 acggaagagagccattgagattgccaatgacacgacttatggcgtgacgtcgtcatt 1260

QY 1261 tatcaaaatctctcaacaggggtctccgtgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1320
Db 1261 tatcaaaatctctcaacaggggtctccgtgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1320

QY 1321 tgcataaactccctttatcccgagacacaaactccgtttggcgcatgaacaaatcg 1380
Db 1321 tgcataaactccctttatcccgagacacaaactccgtttggcgcatgaacaaatcg 1380

QY 1381 ggctcaggcagagatgagcgaagagggcgaagggcgaagggcgaagggcgaagggcgaaggg 1440
Db 1381 ggctcaggcagagatgagcgaagagggcgaagggcgaagggcgaagggcgaagggcgaaggg 1440

QY 1441 aatatccacgtcaacatagatga 1464
Db 1441 aatatccacgtcaacatagatga 1464

RESULT 3

US-09-351-823-4
; Sequence 4, Application US/09351823
; GENERAL INFORMATION:
; APPLICANT: Duivick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fumonisin Detoxification
; FILE REFERENCE: 5718-56
; CURRENT APPLICATION NUMBER: US/09/351,823
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:

QY 361 tactatgagctggccgacaagatcgccggaagaccccttcccgaggacaacggaag 420
Db 361 tactatgagctggccgacaagatcgccggaagaccccttcccgaggacaacggaag 420
QY 421 ccgaattggcgttacgagccgatagggggtgtgtgctggtattgccagctggaacgcgact 480
Db 421 ccgaattggcgttacgagccgatagggggtgtgtgctggtattgccagctggaacgcgact 480
QY 481 ttctttaagctggctggaagatagcccgccctcgcccgccgctctctctcatcttc 540
Db 481 ttctttaagctggctggaagatagcccgccctcgcccgccgctctctctcatcttc 540
QY 541 aaagcctcgagaaaatcccccgtggcggttcttgggcctcgctctctctctcgagaagcc 600
Db 541 aaagcctcgagaaaatcccccgtggcggttcttgggcctcgctctctctcgagaagcc 600
QY 601 ggaattccctcgagctgagctgagcttccctcaactggagcacgagtgacggatg 660
Db 601 ggaattccctcgagctgagctgagcttccctcaactggagcacgagtgacggatg 660
QY 661 ggcgcacatgacattgcgaagatcacgcttcacaagatctgtcgccgctgcccgcgc 720
Db 661 ggcgcacatgacattgcgaagatcacgcttcacaagatctgtcgccgctgcccgcgc 720
QY 721 gtaagcaagcaacatcaagtccaacatgaagcgcgtcactctagaactgggggaaaag 780
Db 721 gtaagcaagcaacatcaagtccaacatgaagcgcgtcactctagaactgggggaaaag 780
QY 781 caaacatcgcttcaagaagctctctcgaaagcagtcgggggaatcgccaaaggaat 840
Db 781 caaacatcgcttcaagaagctctctcgaaagcagtcgggggaatcgccaaaggaat 840
QY 841 ttctcaaaaattcgggcaaaattgggtcccccctcctctgttctagtcaaatgggaaat 900
Db 841 ttctcaaaaattcgggcaaaattgggtcccccctcctctgttctagtcaaatgggaaat 900
QY 901 ttacgagaaaattccatggagtcgctatggctcatatttggaagctgtcagagatggctt 960
Db 901 ttacgagaaaattccatggagtcgctatggctcatatttggaagctgtcagagatggctt 960
QY 961 ggcggaacccattggaaccccaagagacgcatgggtcccttcgtcgacaaagtcaccagtac 1020
Db 961 ggcggaacccattggaaccccaagagacgcatgggtcccttcgtcgacaaagtcaccagtac 1020
QY 1021 gacagatcttgggttaacattgacgttggcaaggatacccgacagctctcactggtgctt 1080
Db 1021 gacagatcttgggttaacattgacgttggcaaggatacccgacagctctcactggtgctt 1080
QY 1081 ggtagaaggcgacaaaggatctgcgattgaaccgagcagatatttgcataatccaaacca 1140
Db 1081 ggtagaaggcgacaaaggatctgcgattgaaccgagcagatatttgcataatccaaacca 1140
QY 1141 ggcagcaaaaattggttgagagatcttggcccgctcttgcatttaagacgttccaag 1200
Db 1141 ggcagcaaaaattggttgagagatcttggcccgctcttgcatttaagacgttccaag 1200
QY 1201 acgaaagaagagccattgagattgccaatgacacgacttatgggctagcctcgttcat 1260
Db 1201 acgaaagaagagccattgagattgccaatgacacgacttatgggctagcctcgttcat 1260
QY 1261 tataccaaatctctcaacagggtctccgtgtctcgtcgccgctcgagacgggtggtc 1320
Db 1261 tataccaaatctctcaacagggtctccgtgtctcgtcgccgctcgagacgggtggtc 1320
QY 1321 tcgatcaacttccctttatcccgagacacaaactcgtttggcggtatgaaacaatcg 1380
Db 1321 tcgatcaacttccctttatcccgagacacaaactcgtttggcggtatgaaacaatcg 1380
QY 1381 ggcctcagagagagctagcggaagaggtcctaagggtacttgagaccaagaccatt 1440
Db 1381 ggcctcagagagagctagcggaagaggtcctaagggtacttgagaccaagaccatt 1440

QY 1441 aatatccacgtcaacatagatga 1464
Db 1441 aatatccacgtcaacatagatga 1464
RESULT 6
US-09-677-682-4
; Sequence 4, Application US/09677682
; GENERAL INFORMATION:
; APPLICANT: Duwick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fummonisin Detoxification
; FILE REFERENCE: 35718/204101 (5718-111B)
; CURRENT APPLICATION NUMBER: US/09/677,682
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: aldehyde dehydrogenase , fully spliced cDNA
US-09-677-682-4
Query Match 100.0%; Score 1464; DB 26; Length 1464;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atggtttcttcctgcacgaatacaagatgaactcttcacaaatgaattcgtctcc 60
Db 1 atggtttcttcctgcacgaatacaagatgaactcttcacaaatgaattcgtctcc 60
QY 61 tccaaggggtccgagagattaaactcaccgaacccgtgggacgaatccaccgttgcact 120
Db 61 tccaaggggtccgagagattaaactcaccgaacccgtgggacgaatccaccgttgcact 120
QY 121 gatgttcacatgggccaacgcccagatgtcagacatgcagtagcgcgttcggtcaggcg 180
Db 121 gatgttcacatgggccaacgcccagatgtcagacatgcagtagcgcgttcggtcaggcg 180
QY 181 gtcaaaaaggcccatggaagaagttcacagggtgcacaacgcggcgctgcatgcttaag 240
Db 181 gtcaaaaaggcccatggaagaagttcacagggtgcacaacgcggcgctgcatgcttaag 240
QY 241 ttccggaacctgcgcgagagaagacgcgagaagctcgtcgtctggtgagctcgtgccacc 300
Db 241 ttccggaacctgcgcgagagaagacgcgagaagctcgtcgtctggtgagctcgtgccacc 300
QY 301 ggtagaccggtgtcgaatgacatcttcgacatttccaaacatggtctcgtgtttcgc 360
Db 301 ggtagaccggtgtcgaatgacatcttcgacatttccaaacatggtctcgtgtttcgc 360
QY 361 tactatgcaagctggccgacaaagatcgccggaagaccttcccgaggacaacggcaag 420
Db 361 tactatgcaagctggccgacaaagatcgccggaagaccttcccgaggacaacggcaag 420
QY 421 ccgaattggcgttacgagccgagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
Db 421 ccgaattggcgttacgagccgagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
QY 481 ttctttaagctggctggaagatagcccgccctcgcccgccgctctctctcatcttc 540
Db 481 ttctttaagctggctggaagatagcccgccctcgcccgccgctctctctcatcttc 540
QY 541 aaagcctcgagaaaatcccccgtggcggttcttgggcctcgctctctctctcgagaagcc 600
Db 541 aaagcctcgagaaaatcccccgtggcggttcttgggcctcgctctctctctcgagaagcc 600
QY 601 ggaattccctcgagctgagctgagcttccctcaactggagcacgagtgacggatg 660

```
Db 601 ggattccctcctggagtcgtgcagttcctcactggagcacgagtgacgggtgaagcattg 660
Qy 661 gcgtcgacatgacatgcgaagatcagcttcaaaatctgtcgcgggtgcgcgcgcc 720
Db 661 ggcgtcgacatgacatgcgaagatcagcttcaaaatctgtcgcgggtgcgcgcgcc 720
Qy 721 gtaagcaagcaactcaagtcacacatgaagcgcgtcactctagaactgggggaaag 780
Db 721 gtaagcaagcaactcaagtcacacatgaagcgcgtcactctagaactgggggaaag 780
Qy 781 ccaacatcgtctcaagcaagctcctcgaacgagtcgagcgagtcggggaagat 840
Db 781 ccaacatcgtctcaagcaagctcctcgaacgagtcgagcgagtcggggaagat 840
Qy 841 ttctcaaaattcgggcaaaattgggtcccccctcctgttgcgtagtgcaatgggaaat 900
Db 841 ttctcaaaattcgggcaaaattgggtcccccctcctgttgcgtagtgcaatgggaaat 900
Qy 901 ttacgagaaattccatggagtcgctcatgctcatttgagagtcgtcagagatggctt 960
Db 901 ttacgagaaattccatggagtcgctcatgctcatttgagagtcgtcagagatggctt 960
Qy 961 ggcagaaacccattggaaccaaaggagcgcgtgctccttcgtcgacaaagtcacagtac 1020
Db 961 ggcagaaacccattggaaccaaaggagcgcgtgctccttcgtcgacaaagtcacagtac 1020
Qy 1021 gacagagtcctgggttaacattgacgttggaagataccgcgcagctcctcactggcgtt 1080
Db 1021 gacagagtcctgggttaacattgacgttggaagataccgcgcagctcctcactggcgtt 1080
Qy 1081 ggtagaagggcgcaagggatctcgatgaacacgacgatatgttcaatcccaaaaca 1140
Db 1081 ggtagaagggcgcaagggatctcgatgaacacgacgatatgttcaatcccaaaaca 1140
Qy 1141 ggcagcaaaattggttgagagatcttggcccgcgtctgttcattaaagacgttcaag 1200
Db 1141 ggcagcaaaattggttgagagatcttggcccgcgtctgttcattaaagacgttcaag 1200
Qy 1201 acggaagaagggccatgaattcgaatcgaacacgacgacttatggctagcctggctatt 1260
Db 1201 acggaagaagggccatgaattcgaatcgaacacgacgacttatggctagcctggctatt 1260
Qy 1261 tatacaaatctctcaacaggggtctcgtgtctcgtggcgctcgagaccggtggcgtc 1320
Db 1261 tatacaaatctctcaacaggggtctcgtgtctcgtggcgctcgagaccggtggcgtc 1320
Qy 1321 tcatcaactctcccttatcccgagacacaaactcgtttggcggtatgaacaaatcg 1380
Db 1321 tcatcaactctcccttatcccgagacacaaactcgtttggcggtatgaacaaatcg 1380
Qy 1381 ggcctcagcagagactaggcgaaggggtcgaaggcgtactttgagcccaagaccatt 1440
Db 1381 ggcctcagcagagactaggcgaaggggtcgaaggcgtactttgagcccaagaccatt 1440
Qy 1441 aatatccagtcacacatagagtga 1464
Db 1441 aatatccagtcacacatagagtga 1464
```

RESULT 7

```
US-09-677-682A-4
; Sequence 4, Application US/09677682A
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 35718/204101
; CURRENT APPLICATION NUMBER: US/09/677,682A
; CURRENT FILING DATE: 2000-10-02
```

```
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; NAME/KEY: misc_feature
; FEATURE:
; LOCATION: (0)...(0)
; OTHER INFORMATION: aldehyde dehydrogenase, fully spliced cDNA
US-09-677-682A-4
```

```
Query Match 100.0%; Score 1464; DB 26; Length 1464;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 atggtttcttcgtcgcgaatacagaagtgaaactcttcatacaaatgaattcgtctcc 60
Db 1 atggtttcttcgtcgcgaatacagaagtgaaactcttcatacaaatgaattcgtctcc 60
Qy 61 tccaaggggtccgagagattaacgctcaagaaacccgtgggacgaatccacggtgccact 120
Db 61 tccaaggggtccgagagattaacgctcaagaaacccgtgggacgaatccacggtgccact 120
Qy 121 gatgttcacgtggcacaacgcgcgcgtatcgcagtcagtcagtcgcttcggtgcagcgc 180
Db 121 gatgttcacgtggcacaacgcgcgcgtatcgcagtcagtcagtcgcttcggtgcagcgc 180
Qy 181 gtcaaaaaggcccatggaagaagtctacaggttcacaaacgcgcgcgtgcgtgttaag 240
Db 181 gtcaaaaaggcccatggaagaagtctacaggttcacaaacgcgcgcgtgcgtgttaag 240
Qy 241 ttcgcggacctcgcgcgaagaacgcgcgcgtcgcgtcgttcggtgcgtgcgcgcacc 300
Db 241 ttcgcggacctcgcgcgaagaacgcgcgcgtcgcgtcgttcggtgcgtgcgcgcacc 300
Qy 301 ggtgaacgggtgtcgtatcactcactcatttcgacattccaaacatggttcggtgttcgc 360
Db 301 ggtgaacgggtgtcgtatcactcactcatttcgacattccaaacatggttcggtgttcgc 360
Qy 361 tactatgcaggctggcgacaaagatcccgaaagaccttcccgagacaaacgcgcgcag 420
Db 361 tactatgcaggctggcgacaaagatcccgaaagaccttcccgagacaaacgcgcgcag 420
Qy 421 ccgaattggcgttacgagccgatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
Db 421 ccgaattggcgttacgagccgatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
Qy 481 ttcttaccgtcgcgtggaagatagccccccctcgcgcgcgcgtcgtccttcctatcttc 540
Db 481 ttcttaccgtcgcgtggaagatagccccccctcgcgcgcgcgtcgtccttcctatcttc 540
Qy 541 aaagcctcggagaaatccccctcgcgcgcgtcgtccttccttccttccttccttccttccttc 600
Db 541 aaagcctcggagaaatccccctcgcgcgcgtcgtccttccttccttccttccttccttccttc 600
Qy 601 ggattccctcctggagtcgtgcagttcctcactggagcacgagtgacgggtgaagcattg 660
Db 601 ggattccctcctggagtcgtgcagttcctcactggagcacgagtgacgggtgaagcattg 660
Qy 661 gcgtcgacatggacattgcgaagatcagcttcaaaatctgtcgcgggtgcgcgcgcc 720
Db 661 gcgtcgacatggacattgcgaagatcagcttcaaaatctgtcgcgggtgcgcgcgcc 720
Qy 721 gtaagcaagcaactcaagtcacacatgaagcgcgtcactctagaactgggggaaag 780
Db 721 gtaagcaagcaactcaagtcacacatgaagcgcgtcactctagaactgggggaaag 780
Qy 781 ccaaccatcgtctcaacgaagctcctcgaacgagtcgagcgagtcgggggaatcggcaagat 840
Db 781 ccaaccatcgtctcaacgaagctcctcgaacgagtcgagcgagtcgggggaatcggcaagat 840
```

Db 781 ccaacacatgcttcaacagaaagctcctctctcgaaacgagcgagtcgagggaatcgcaaaagat 840
Qy 841 ttctcaaaattcgggcaaaatttggttccccccctcctgtttctagtgcaatgggaaat 900
Db 841 ttctcaaaattcgggcaaaatttggttccccccctcctgtttctagtgcaatgggaaat 900
Qy 901 tttagcggagaaattccatggagtcgctcatggctcatttggagctgtcagagatggctt 960
Db 901 tttagcggagaaattccatggagtcgctcatggctcatttggagctgtcagagatggctt 960
Qy 961 ggcagaacccattgaacccaagagacagatggtccctctgctcgacaagtccagatc 1020
Db 961 ggcagaacccattggaacccaagagacagatggtccctctgctcgacaagtccagatc 1020
Qy 1021 gacagagcttgggttaacattgacgttggcaaggatcccgacgacgtccctcaactggcgtt 1080
Db 1021 gacagagcttgggttaacattgacgttggcaaggatcccgacgacgtccctcaactggcgtt 1080
Qy 1081 ggtagaaaggcgacaaaggattcgcgattgaaccgacgatatgttcaatcccaaacca 1140
Db 1081 ggtagaaaggcgacaaaggattcgcgattgaaccgacgatatgttcaatcccaaacca 1140
Qy 1141 ggcagcaaaatttggtttagagagatctttggccccctcttgcatttaagagcttcaag 1200
Db 1141 ggcagcaaaatttggtttagagagatctttggccccctcttgcatttaagagagcttcaag 1200
Qy 1201 acggaagagagccattgagattgccaatgacagacattatgggctagcctcggtcatt 1260
Db 1201 acggaagagagccattgagattgccaatgacagacattatgggctagcctcggtcatt 1260
Qy 1261 tataccaatctcacaagggttccggtgtctgtctgtctgtctgtctgtctgtctgtctgtc 1320
Db 1261 tataccaatctcacaagggttccggtgtctgtctgtctgtctgtctgtctgtctgtctgtc 1320
Qy 1321 tcatcaacttcccttttcccgagacacaaactccgtttgtggcgcatgaacaactcg 1380
Db 1321 tcatcaacttcccttttcccgagacacaaactccgtttgtggcgcatgaacaactcg 1380
Qy 1381 ggcctaggcagagactagcgaagaaggctcaaggctacttggagcccaagaccatt 1440
Db 1381 ggcctaggcagagactagcgaagaaggctcaaggctacttggagcccaagaccatt 1440
Qy 1441 aatatccacgtcaacatagatga 1464
Db 1441 aatatccacgtcaacatagatga 1464

RESULT 8

US-09-677-682B-4
; Sequence 4, Application US/09677682B
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/204101
; CURRENT APPLICATION NUMBER: US/09/677,682B
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: aldehyde dehydrogenase, fully spliced cDNA

US-09-677-682B-4

Query Match 100.0%; Score 1464; DB 26; Length 1464;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgggtctcttcgctgacgaatacacaagagtgaaactcttcatcaacaatgaattcgtctcc 60
Db 1 atgggtctcttcgctgacgaatacacaagagtgaaactcttcatcaacaatgaattcgtctcc 60
Qy 61 tccaaggggtccgagagattaaacgctcacgaaccggtggagcgaatccaccgcttgcacct 120
Db 61 tccaaggggtccgagagattaaacgctcacgaaccggtggagcgaatccaccgcttgcacct 120
Qy 121 gatgttcaactggtggccaaacgcccagatgtcgacagtgcagtagcgcgttcggtgcagggcg 180
Db 121 gatgttcaactggtggccaaacgcccagatgtcgacagtgcagtagcgcgttcggtgcagggcg 180
Qy 181 gtcaaaaggccctcgtggaagaattcacaggttcacaggtgcacaacgcgcgcgtgcattgaag 240
Db 181 gtcaaaaggccctcgtggaagaattcacaggttcacaggtgcacaacgcgcgcgtgcattgaag 240
Qy 241 ttccggaacctcgcgcgagaagaaccccgagaagctcgcgtctctggagtcgctgcccacc 300
Db 241 ttccggaacctcgcgcgagaagaaccccgagaagctcgcgtctctggagtcgctgcccacc 300
Qy 301 ggtagaccggtgtcgatgatacactcatttcacataccaaacatggttcctcgttcttcgc 360
Db 301 ggtagaccggtgtcgatgatacactcatttcacataccaaacatggttcctcgttcttcgc 360
Qy 361 tactatcgaggctggccgacaagaatcgccgaaagacctttcccgagagacacagcaag 420
Db 361 tactatcgaggctggccgacaagaatcgccgaaagacctttcccgagagacacagcaag 420
Qy 421 ccgaattggcgtttacgagccgattgggggtgtgtgctggtattgccagctgggaacgcgact 480
Db 421 ccgaattggcgtttacgagccgattgggggtgtgtgctggtattgccagctgggaacgcgact 480
Qy 481 ttctttcacatcgctggaagatagcccccgccctcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
Db 481 ttctttcacatcgctggaagatagcccccgccctcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
Qy 541 aaagcctcggagaaatccccgcgtggcggtttctgggcctcgcctctctctctcgcagaagcc 600
Db 541 aaagcctcggagaaatccccgcgtggcggtttctgggcctcgcctctctctctcgcagaagcc 600
Qy 601 ggattccctcctggagtcgtgcagttcctcactggagacacagtgacgggttgaaagcattg 660
Db 601 ggattccctcctggagtcgtgcagttcctcactggagacacagtgacgggttgaaagcattg 660
Qy 661 gcgtgcacatggacattcgaaagatcagcttcaagaatctgctgcgcgcgcgcgcgcgcgcgcgc 720
Db 661 gcgtgcacatggacattcgaaagatcagcttcaagaatctgctgcgcgcgcgcgcgcgcgcgcgc 720
Qy 721 gtaagcaagcaacactcaagtccaactgaagcgcgtcactctagacttagactgggggaaag 780
Db 721 gtaagcaagcaacactcaagtccaactgaagcgcgtcactctagacttagactgggggaaag 780
Qy 781 ccaacacatcgcttcaacgaagctcctcgcgaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 840
Db 781 ccaacacatcgcttcaacgaagctcctcgcgaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 840
Qy 841 ttctcaaaattcgggcaaaatttggttccccccctcctctgttctagtcgaatggggaat 900
Db 841 ttctcaaaattcgggcaaaatttggttccccccctcctctgttctagtcgaatggggaat 900
Qy 901 tttagcggagaaattccatggagtcgctcatgttggagcttggagctgcagagatggctt 960
Db 901 tttagcggagaaattccatggagtcgctcatgttggagcttggagctgcagagatggctt 960
Qy 961 ggcagaacccattgaacccaagagacagatggtccctctgctcgacaagtccagatc 1020
Db 961 ggcagaacccattgaacccaagagacagatggtccctctgctcgacaagtccagatc 1020

Db 961 ggcagaaacccattggaaccccaaggagacgcatggtccctctgtcgacaaagtccacgtac 1020
Qy 1021 gacagagtttggttaacattgacgttggcaaggataccgacgactctcactggogtt 1080
Db 1021 gacagagtttggttaacattgacgttggcaaggataccgacgactctcactggogtt 1080
Qy 1081 ggtagaaaggcgacaaaggattcgcgattgaaccgacgatatattgtcaatcccaaaaca 1140
Db 1081 ggtagaaaggcgacaaaggattcgcgattgaaccgacgatatattgtcaatcccaaaaca 1140
Qy 1141 ggcagcaaaatttggttgagagatctttggcccccgctctgtccattaaagcgttcaag 1200
Db 1141 ggcagcaaaatttggttgagagatctttggcccccgctctgtccattaaagcgttcaag 1200
Qy 1201 acggaagaagagccattgagattgccaatgacgacttatggctagctcgttcatt 1260
Db 1201 acggaagaagagccattgagattgccaatgacgacttatggctagctcgttcatt 1260
Qy 1261 tatacaaatctctcaacagggtctcgtgtctctgtggcgctcgagaccggtggcgctc 1320
Db 1261 tatacaaatctctcaacagggtctcgtgtctctgtggcgctcgagaccggtggcgctc 1320
Qy 1321 tcgatacaattccctttatcccgagacaaactccgcttggcgcatgaaacaactcg 1380
Db 1321 tcgatacaattccctttatcccgagacaaactccgcttggcgcatgaaacaactcg 1380
Qy 1381 ggcctggcagagctagtcgaaagggtcctcaaggcgctacttgagcccaagaccatt 1440
Db 1381 ggcctggcagagctagtcgaaagggtcctcaaggcgctacttgagcccaagaccatt 1440
Qy 1441 aatatccagctcaacatagatga 1464
Db 1441 aatatccagctcaacatagatga 1464

RESULT 9

US-09-882-694-4
; Sequence 4: Application US/09882694
; GENERAL INFORMATION:
; APPLICANT: Duwick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 35718/208255
; CURRENT APPLICATION NUMBER: US/09/882,694
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: aldehyde dehydrogenase, fully spliced cDNA
US-09-882-694-4

Query Match 100.0%; Score 1464; DB 33; Length 1464;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atggtttcttcgctgacgaatacaagagtgaaactttcatcaacaatgaattcgtctcc 60
Db 1 atggtttcttcgctgacgaatacaagagtgaaactttcatcaacaatgaattcgtctcc 60
Qy 61 tccaagggttcgagagattaaacgtccagaaacccggtgggacgaatccaccggttgcact 120

Db 61 tccaagggttcgagagattaaacgtccagaaacccggtgggacgaatccaccggttgcact 120
Qy 121 gatgttcacgtggccaaacgcgcgcgatgtcgacagtgacgagtcgcggttcggtcagggc 180
Db 121 gatgttcacgtggccaaacgcgcgcgatgtcgacagtgacgagtcgcggttcggtcagggc 180
Qy 181 gtcaaaaaaggcccatggaagaagttcacaggtgcacaacgcgcgcggtgcgttcgttaag 240
Db 181 gtcaaaaaaggcccatggaagaagttcacaggtgcacaacgcgcgcggtgcgttcgttaag 240
Qy 241 ttccgagaccctcccgagaaagcgcgcgagagtcgctcgctgagtcgctgcgccacc 300
Db 241 ttccgagaccctcccgagaaagcgcgcgagagtcgctcgctgagtcgctgcgccacc 300
Qy 301 ggttagaccggttcgagatcactcatttcgacattccaaacatggttcctcgtgttcgc 360
Db 301 ggttagaccggttcgagatcactcatttcgacattccaaacatggttcctcgtgttcgc 360
Qy 361 tactatgcaggtggccgacaaagatcgcgcgaaagacaccttcccgagagacaacggcaag 420
Db 361 tactatgcaggtggccgacaaagatcgcgcgaaagacaccttcccgagagacaacggcaag 420
Qy 421 ccgaattggcgttacgagccgatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
Db 421 ccgaattggcgttacgagccgatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
Qy 481 ttctttacgtcggctggaagatagcccgcccgcccgcccgcccgcccgcccgcccgcccgcc 540
Db 481 ttctttacgtcggctggaagatagcccgcccgcccgcccgcccgcccgcccgcccgcccgcc 540
Qy 541 aaagcctcggagaaatcccccgtggcggttctctgggctcgtcctctctctcgcagaagcc 600
Db 541 aaagcctcggagaaatcccccgtggcggttctctgggctcgtcctctctctcgcagaagcc 600
Qy 601 ggattccctcctggagtcgtgcagttccctcaactggagcacgagtcagcgggtgaagcattg 660
Db 601 ggattccctcctggagtcgtgcagttccctcaactggagcacgagtcagcgggtgaagcattg 660
Qy 661 gcgtcgacatgacattgcgaagatcagcttcacaagatctgtcgcggtggcgcgcc 720
Db 661 gcgtcgacatgacattgcgaagatcagcttcacaagatctgtcgcggtggcgcgcc 720
Qy 721 gtcaagcaagcaacactcaagtcacaatgaagcgcgtcactctagaactgggggaaag 780
Db 721 gtcaagcaagcaacactcaagtcacaatgaagcgcgtcactctagaactgggggaaag 780
Qy 781 ccaaccatcgtcttcaacgagctcctctcgaaacgcgagtcgagggtgcaaaaggat 840
Db 781 ccaaccatcgtcttcaacgagctcctctcgaaacgcgagtcgagggtgcaaaaggat 840
Qy 841 ttctcaaaattcgggcaaaatttggttccccccctcctgtttgtagtcaatgggggaaat 900
Db 841 ttctcaaaattcgggcaaaatttggttccccccctcctgtttgtagtcaatgggggaaat 900
Qy 901 tttagcggagaaattccatggagtcgctcattggcctatttgagggtcgtcagagatggctt 960
Db 901 tttagcggagaaattccatggagtcgctcattggcctatttgagggtcgtcagagatggctt 960
Qy 961 ggccagaaccattggaaacccaagagacgcatggtcccttcgctcagcaagtcaccagtac 1020
Db 961 ggccagaaccattggaaacccaagagacgcatggtcccttcgctcagcaagtcaccagtac 1020
Qy 1021 gacagagtttggttaacattgacgttggcaaggataccgacgactctcactggcggtt 1080
Db 1021 gacagagtttggttaacattgacgttggcaaggataccgacgactctcactggcggtt 1080
Qy 1081 ggtagaaggcgacaaaggattcgcgattgaaccgacgatatattgtcaatcccaaaaca 1140
Db 1081 ggtagaaggcgacaaaggattcgcgattgaaccgacgatatattgtcaatcccaaaaca 1140
Qy 1141 ggcagcaaaatttggttgagagatctttggcccccgctctgtccattaaagcgttcaag 1200

Db 1141 ggcagcaaaatttggttgaggagatcttggcccgctcttgctcattaaagagcttaag 1200
Qy 1201 acggaagaagagccattgagatgccaatgacacgaccttatggctagctcggtcatt 1260
Db 1201 acggaagaagagccattgagatgccaatgacacgaccttatggctagctcggtcatt 1260
Qy 1261 tataccaaatctctcaacaggggtctccgtgtctctgtggtggtggtggtggtggtggtc 1320
Db 1261 tataccaaatctctcaacaggggtctccgtgtctctgtggtggtggtggtggtggtc 1320
Qy 1321 togataactctcccttatcccgagagacacaaactccgtttggtggtggtggtggtggtc 1380
Db 1321 togataactctcccttatcccgagagacacaaactccgtttggtggtggtggtggtggtc 1380
Qy 1381 ggctcaggcagagagctaggcggaagggtctcaagggtctacttgagcccaagaccatt 1440
Db 1381 ggctcaggcagagagctaggcggaagggtctcaagggtctacttgagcccaagaccatt 1440
Qy 1441 aatatcccgctcaacatagagtga 1464
Db 1441 aatatcccgctcaacatagagtga 1464

RESULT 10

US-09-882-694A-4
; Sequence 4, Application US/09882694A
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/882,694A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala splnifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: aldehyde dehydrogenase, fully spliced cDNA
US-09-882-694A-4

Query Match 100.0%; Score 1464; DB 33; Length 1464;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 atggtctcttccttcgcaatacaagatgaactcttcatacaaatgaattgctctcc 60
Db 1 atggtctcttccttcgcaatacaagatgaactcttcatacaaatgaattgctctcc 60
Qy 61 tccaagggttcgagagattaacgtctacgaaccctgtgggacgaatccaccgttgccact 120
Db 61 tccaagggttcgagagattaacgtctacgaaccctgtgggacgaatccaccgttgccact 120
Qy 121 gatgttcagtaggccaacggtcgatgtcgacagtgcagtagccgttcggtcaggcg 180
Db 121 gatgttcagtaggccaacggtcgatgtcgacagtgcagtagccgttcggtcaggcg 180
Qy 181 gtcaaaagggcccatggaagaagttcacaggtgcacaacgcggtgcatgtcttaag 240
Db 181 gtcaaaagggcccatggaagaagttcacaggtgcacaacgcggtgcatgtcttaag 240
Qy 241 ttcgggagacctcgcgagagaagacgcgagagctcgtctgtgagctcgtccacc 300

Db 241 ttcgggagacctcgcgagagaagacgcgagagctcgtctgtgagctcgtccacc 300
Qy 301 ggtagaccggtgtcgatgatactcatttcgacattccaaacatggtctcagtttgcgc 360
Db 301 ggtagaccggtgtcgatgatactcatttcgacattccaaacatggtctcagtttgcgc 360
Qy 361 tactatgcaggctgggcccgaagaatgcgcggaagaccttcccgaggaacaacggcaag 420
Db 361 tactatgcaggctgggcccgaagaatgcgcggaagaccttcccgaggaacaacggcaag 420
Qy 421 ccgaattggcgtttacgagccggtgggtgtgtgtggtgtgtgtggtgtgtgtggtgtgt 480
Db 421 ccgaattggcgtttacgagccggtgggtgtgtgtggtgtgtgtggtgtgtgtggtgtgt 480
Qy 481 tttctttacgtcgcgtggaagatagccccccctcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
Db 481 tttctttacgtcgcgtggaagatagccccccctcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
Qy 541 aaagcctcggagaaatcccgctggggttctgggctcgtctctctctctctctctctctctct 600
Db 541 aaagcctcggagaaatcccgctggggttctgggctcgtctctctctctctctctctctctct 600
Qy 601 ggattccctcgtcagtcgttcagttcctcactgagacgacgagtgacgggtggaagcatt 660
Db 601 ggattccctcgtcagtcgttcagttcctcactgagacgagtgacgggtggaagcatt 660
Qy 661 gctcgcacatggacattgcaagatcagcttcaagatctgtcggcgtgtggtcgcgcgcgc 720
Db 661 gctcgcacatggacattgcaagatcagcttcaagatctgtcggcgtgtggtcgcgcgcgc 720
Qy 721 gtcagcaagcaacactcaagtcacacatgagcgtcactctagacactgaggggaaag 780
Db 721 gtcagcaagcaacactcaagtcacacatgagcgtcactctagacactgaggggaaag 780
Qy 781 ccaaccatcgtcttcaacgaagctcctctcgaacggcagtcggggggaatcgccaagagat 840
Db 781 ccaaccatcgtcttcaacgaagctcctctcgaacggcagtcggggggaatcgccaagagat 840
Qy 841 tctcaaaaattcgggcaaaattgggtccccccctcctctgtgtgtgtgtgtgtgtgtgtgt 900
Db 841 tctcaaaaattcgggcaaaattgggtccccccctcctctgtgtgtgtgtgtgtgtgtgtgt 900
Qy 901 ttacggagaaattccatggagtcgttcagtcgttcattttggaggtctcagagatggctt 960
Db 901 ttacggagaaattccatggagtcgttcagtcgttcattttggaggtctcagagatggctt 960
Qy 961 ggcagaaaccttggaaccgaagacgacgacgacgacgacgacgacgacgacgacgacgac 1020
Db 961 ggcagaaaccttggaaccgaagacgacgacgacgacgacgacgacgacgacgacgacgac 1020
Qy 1021 gacagagctctgggttaacattgacgttggcagagataccgagcagctcactggcgtt 1080
Db 1021 gacagagctctgggttaacattgacgttggcagagataccgagcagctcactggcgtt 1080
Qy 1081 ggtagaagggcgcaacagggattcgcgattgacccgacgacgacgacgacgacgacgacgac 1140
Db 1081 ggtagaagggcgcaacagggattcgcgattgacccgacgacgacgacgacgacgacgacgac 1140
Qy 1141 ggcagcaaaatttgggttgaggagatctttggccccctccttgccttgccttaagacgttcaag 1200
Db 1141 ggcagcaaaatttgggttgaggagatctttggccccctccttgccttgccttaagacgttcaag 1200
Qy 1201 acggaagaagggccattgagattgccaatgacacgaccttatggctagctcgtggtcatt 1260
Db 1201 acggaagaagggccattgagattgccaatgacacgaccttatggctagctcgtggtcatt 1260
Qy 1261 tataccaaatctctcaacaggggtctccgtgtctctgtggtggtggtggtggtggtggtc 1320
Db 1261 tataccaaatctctcaacaggggtctccgtgtctctgtggtggtggtggtggtggtggtc 1320
Qy 1321 tcgatacactccctcttatcccgagagacacaaactccgtttggtggtggtggtggtggtc 1380
Db 1321 tcgatacactccctcttatcccgagagacacaaactccgtttggtggtggtggtggtggtc 1380

Db 1321 tcgatacaactccctttatccccgagacacaaactccggtttggtgagcgcatgaaacaatcg 1380
Qy 1381 ggtcaagcagagagctaggcgaagaggtctcaagcgtaactttagagcccaagaccatt 1440
Db 1381 ggtcagcagagagctaggcgaagaggtctcaagcgtaactttagagcccaagaccatt 1440
Qy 1441 aatatccacgtcaacatagagtga 1464
Db 1441 aatatccacgtcaacatagagtga 1464

RESULT 11

US-08-612-853-1

; Sequence 1, Application US/08612853

; GENERAL INFORMATION:

; APPLICANT: Achatz, Gernot

; APPLICANT: Oberkofler, Hannes

; APPLICANT: Simon, Birgit

; APPLICANT: Unger, Andrea

; APPLICANT: Lechenauer, Erich

; APPLICANT: Hirschwehr, Reinhold

; TITLE OF INVENTION: Recombinant Cladosporium Herbarum

; TITLE OF INVENTION: Allergens

; NUMBER OF SEQUENCES: 117

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds LLP

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/612,853

; FILING DATE: 26-FEB-1996

; CLASSIFICATION: 536

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: PCT/AT94/00120

; FILING DATE: 02-MAR-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Friebe, Thomas E

; REGISTRATION NUMBER: 29,258

; REFERENCE/DOCKET NUMBER: 6530-021-999

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-790-9090

; TELEFAX: 212-869-9741

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1491 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA to mRNA

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Cladosporium herbarum

; DEVELOPMENTAL STAGE: Spores and vegetative hyphae

; FEATURE:

; NAME/KEY: Coding Sequence

; LOCATION: 1...1488

; OTHER INFORMATION:

US-08-612-853-1

Query Match 18.3%; Score 267.6; DB 10; Length 1491;
Best Local Similarity 53.0%; Pred. No. 3.1e-63;
Matches 766; Conservative 0; Mismatches 654; Indels 26; Gaps 8;

Qy 17 aqgaatacaagagtgaaactcttcatcaacaatgaattcgctctctcaagggtccgaga 76
Db 41 ACAGACGCCGACCGCCCTCTTCAACAACAGAGTTCGTCACAGGGCCACAGAGGCAAGA 100
Qy 77 gattaacgctcacgaacccgtgggacgaatccacccgttggccactgatgttcaagtgccca 136
Db 101 CCTTCGATGTCATCAACCCCTCCGACGA---GAGCGTGATCACCCAGGTCCACGAGGCCA 157
Qy 137 acgcgccgatgtcgacagtgacgtcggttcggttcgagcggttcgaaagagggcccat 196
Db 158 CCGAGAAAGATGTCGACATCGCGTCGCGGCCGCCCAAGCCCTTC---GAGGGCTCAT 214
Qy 197 ggaagaagttcacaggtgcacaaacgcgcggcggtgcttaagtttcgcggagaccccgccy 256
Db 215 GGAGACTGGAGACACCCGAGAACCGTGGCAAGCTGCTCAACAACCTCGCCAACTGTGTG 274
Qy 257 aqaagaaccccgagaagctcgctgctgtggagtcgctgcccacacggttagaccggtgctga 316
Db 275 AGAAGAACACTGACCTCCTTGTGCGCGTTGAGTCGCTGCACACGCGCAAGGCCACTTCCA 334
Qy 317 tgataactcatttcgacattccaaacatggctcogtctgttgcctactatgcaggtggg 376
Db 335 TG---GCAAGGTGACATCAGCATGTGCGTCCGGCTGCCTCAGATACTACGGGTGGTGG 391
Qy 377 cgcacaagatccgcggaagaccttcccgaggaacacgcgcaagccgaattggcgcttac- 435
Db 392 CGGACAAGATCACCGCAAGGTCAACGACACTAGCCCGACACTTCAACTACGTCACAAGA 451
Qy 436 --gacccgagtggtgtgctgtgatttgcagctggaacgcgaactttcttaccgtcg 493
Db 452 AGGAGCCCATTTGTTGCCGCTCAGATCATTCCTGGAACTTCCCTCTCTCATGTGGG 511
Qy 494 gctggagaatagccccccctcgcgcggctgctctctcatcttcaaaagcctcggaga 553
Db 512 CATGGAAGATCGGCCCGGCCATTGCTTGGGTAACTGCTGCTCTGAAGACTGCTGAGC 571
Qy 554 aatcccccgctggcggttcctggcgctcctctcttcgcagaagcggatctccctcctg 613
Db 572 AGACCCCTCTTGGTGGTCTGCTGCTGCGTCAGCCCTCGTCAAGGAGCGCGTTTCCCTCCT 631
Qy 614 gagtcgagttcctcactgagacacga---gtgacgggtggaagcattggcgctcgaca 670
Db 632 GTGTCATCAACGTCAATTTCCCGTTCGGCAAGGTGCGTGGTGGCTCTCTCTTCTCACA 691
Qy 671 tgacatttcggaagatcagcttcaacagatctgtcggcggtggcgcgcgctcgaagcaag 730
Db 692 TGGACGTGACAAAGGTGGGCTTCCACCGTTCCACCGTTGTCGCGCGCACAAATCCTCAAG 751
Qy 731 caacactcaagtcacaacatgaagcgcgtcactcttagaactgggggaaagcccaacatcg 790
Db 752 CTGCTGCCCTTCCAACTTGAAGAAGGTCAACCTCGAGCTCGGTGGCAAGTCAACCAACA 811
Qy 791 tcttcaacgaagctcctcctcgaaacgcgcggtcggggaatcggaagatttctcaaat 850
Db 812 TTGTCCTTCGAGGAGCGCGATATTGACAACGGCATCTCATGGTCAACTTCGGTATCTTCT 871
Qy 851 tggggcaaatgtgggtcccccctctctgttagtgcgaatgggaaattttagcgaga 910
Db 872 TCAACCCAGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 924
Qy 911 aattccatggagtcgctcatttggtcatttgag-gctgacagataggttggcagaaac 969
Db 925 ATCTACGACAAGTTGCTCCAGAAAGTCAAGGAGCGCGCACAGAGAAGACGTTGTTGGCGAC 984
Qy 970 catttggaacccaagagagcgcgtgctccctcgttcgacaagtcgccagtagccagagtc 1029
Db 985 CCTTCGCGCGCGACACATTCCAGGTCCTCAGGTTTCCAGGTTTTCAGTTTCAGTTCACGCA 1044
Qy 1030 ttgggttaacattgacgttggcaagataacc---ggcgagctcctcactgagcgttggtaga 1086
Db 1045 ATGGAGTACATCCAGGCCCGCAAGGACGCGGTGCTGCCACCGTCGAGACCGGTGGAAGCGGT 1104

```
QY 1087 aagggcgacgaaggattccgattgaacgaagacgatatttgcataccccaaaccaggaac 1146
DB 1105 AAGGGTGACAAAGGGCTACTTCAATTGAGCCACCACCATCTTCCAACTCACAGAGACATG 1164
QY 1147 aaaaatttggttgaagagatcttggcccgcttctgtccattaaagacgttcaagacgaa 1206
DB 1165 AAGATCGTGAAGAGGAGATCTTCGGCCCGCTCTGCTCCATGCCCAAGTTCAGACCAAG 1224
QY 1207 gaagagccattagattgccaatgacgaagacttattgggtgactcgctgctatttataacc 1266
DB 1225 GAGGACGCCATCAAGCTCGGCAACGCCACATACGGCTCTGCGCGCCGCTCCACACC 1284
QY 1267 aaatctctcaacaggggtctccgtgtctgtcggtcggtcgagaccggtggcgtctcgatc 1326
DB 1285 AAGAACCTCAACACCGCATCGAGGTCTCCAAAGCTCTCAAGCGGCGACCGTCTGGGTC 1344
QY 1327 aacttcccccttattccccgagacacaaaactccgtttggcggtatgaacaaatcgggctca 1386
DB 1345 AACACTTACAAACACCTCCACACACAGATGCCGTTCCGGTGGCTACANGAGTCCGGTATT 1404
QY 1387 ggcagagctagcgaagggctcaaggcgacttggagcccaagacacattaatatc 1446
DB 1405 GCGCGTAGCTCGCGAGGATCGTGGCCAACTACACTCAGACCAAGACTGTTAGCATT 1464
QY 1447 cagctc 1452
DB 1465 CGTCTC 1470

RESULT 12
US-09-404-520-27986
; Sequence 27986, Application US/0904520
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Ghodssi, Azita
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: McIninch, James
; APPLICANT: Timberlake, William E.
; APPLICANT: Yu, Jaehyuk
; TITLE OF INVENTION: EmERICELLA NIDULANS Genome Sequence and Uses Thereof
; FILE REFERENCE: 38-10(15498)A
; CURRENT APPLICATION NUMBER: US/09/404,520
; CURRENT FILING DATE: 1999-09-23
; NUMBER OF SEQ ID NOS: 44345
; SEQ ID NO 27986
; LENGTH: 1491
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-09-404-520-27986
```

```
Query Match 17.1%; Score 250.6; DB 18; Length 1491;
Best Local Similarity 52.5%; Pred. No. 1.7e-58;
Matches 753; Conservative 0; Mismatches 644; Indels 36; Gaps 8;

QY 34 ctcttcatacaaatgaattcgtctctctccaaagggttcgagagatttaacgtctacgaac 93
DB 58 ctgtctatcaacaacagattctcgtaaggcggttgagggcaagacattccaggctcatcaac 117
QY 94 ccgtggagcaaatccacggttgcacatgatgttcacgttgccaaacggtcggtatctgac 153
DB 118 cctctccacga--gaaggtcatcactccgtccacgaagccacgagagagattga- 173
QY 154 agtcagatagccgttcgtgagcggtcctcaaaaaggcccatggaagaagtccacaggt 213
DB 174 --tgcgcccgtctgctgcgcgtctgttggggcccatggcgccaggttcacccc 231
QY 214 gcaacagcgcggtgcgtgttaagttcgcggacctcgcgagaaagcgcgagag 273
DB 232 tctgagcgtggcattttgatcaacaagctggcggtatctgattgagcgtgatcgacac 291
QY 274 ctgcgtcgttgagtcgtgcccacggtgacccggtgctgattgacattcatttcgac 333
```

```
DB 292 ctgcgcgtctatcgagtcctctgcacaacggcaagggttttccattaccattggcca--aggtcgat 348
QY 334 attccaaaacatggtctccgctgtttcgtactactgagcgctggcgacgaagatgcgcgga 393
DB 349 ctgcgaactccatggttgcgtactactacgctggtggtggcggaacagattccacggt 408
QY 394 aagaactttcccgagaggaacaaaggcgaagccgaattgg--cgttacgagcccgatgggggtg 450
DB 409 cagaccattgacccaaccccccgagactcttacctaacocccgacgaagccgcttgggtgt 468
QY 451 tgtgctgggtattgcccagctgggaacgcgacttttcttacctgcggtgggaagatagccccc 510
DB 469 tgcggtcagatcatccccctgggaactccccctcttgatgtggtcctggagatggaaccc 528
QY 511 gccctcgcccgcggtgcgtccctctatcttcaaaagcctcgagaaatccccctgggggtt 570
DB 529 gctgttgcgcgtggtaaacactgtgtccctcaagaccgcccgacagaccctctctccgccc 588
QY 571 ctgggctcgctcctctctcttcgcagaagccgggattccctctcctggagtcgtgcagttcctc 630
DB 589 ctttaacgctgctaagctgatacgaagggtccattccccgcgtggtgatacaacgctcatc 648
QY 631 actgg--agcacgagtgacgggtgaagcattggcgctgcgcacatgacatgacatgcaagaatc 687
DB 649 tctggcttggccgtacccgctggtgctgccatctccagccacacatgacatgacaaggtt 708
QY 688 agottcacaagatctgtcggcggtggccgcgcgttaagcagaagaacactcaagtccaac 747
DB 709 gccttcaactggcttactctgttggaactaactctcctgcaggccgctgtctaaagcaac 768
QY 748 atgaagcgcgtcactctagaaactgggggaaagccaaacatcgtcttcaacgaagctcct 807
DB 769 ctgaagaaggtcaactcttgatcgtcgtggcaagtctcccaacatcgtcttctgagtgcct 828
QY 808 ctogaacggcgagtcgggggaatcgccaagaagat-----ttctcaaaaattcgggcaaat 861
DB 829 gacattgacaacgccaattctccctggccaaacttctgtatctcttcaacccagccagtcg 880
QY 862 tgggttccccctctctgttgcagtgcgaatgggggaaatttagcggagaaattccatgga 921
DB 889 tgcgtgctggatcccgatactcctggtccagggagggcatctacgacaagtctcgtgcgcgc 948
QY 922 gtccgtcatggtcatttggaggtgctcagagatggtgtgccagaaacccatttgaagccc 981
DB 949 ttcaaggagcgtgc-----ccagaagaacaaggctcggaaccccttcgagcag 996
QY 982 aagaggacgcagtggtccctctgcgacaagtcccgatgcgacagatcttgggttaacatt 1041
DB 997 gacaccttcaggggtcccccaggtttcccaagctccagttccgacctatcatgtagtcatc 1056
QY 1042 gacgttggc--aaggataccgcgcagctcctcactggcgttggtagaagggcgacaag 1098
DB 1057 aaccacggcaagaaggctggtgtaccgtgcgcacccggtggtgaccgccacggcaacgag 1116
QY 1099 ggaattcgagatgaaacccgacgatatatttgcatacccaaacccagcagcaaaatttgggtt 1158
DB 1117 ggttacttcatccagccctactgtcttcaacacgctcaactctccgacatgaagatgcccag 1176
QY 1159 gaggagatcttggcccgctgttgcattaaagacgttcaagacgggaggaagagggcatt 1218
DB 1177 gaggagatctcgttccctgctcactatccagaagttccaaaggtggtggtgagggctatc 1236
QY 1219 gagattgcaatgacagcacttatgggtgacctcgtcatttatatacaaatctctcaac 1278
DB 1237 agatcggaactcgacgcagctacgggtcttctgctgctgacacaaagaacagctcaac 1296
QY 1279 aggggtctcgtgctgcgtcgcgcgtcgcagaccgggtggcgtctcgcatacaacttccccctt 1338
DB 1297 accgcattctcgtgttccaaacgctcgaagctggaagctggaacgctcgtggaatacaacac 1356
QY 1339 attccccgagacacaaactccgtttggcgcatgaaacaaatcgggtcagggcagagagcta 1398
DB 1357 atgatctgtaccaggtccctcctcgttggcttcaagcagtcggtctcgcgcgcgttgcgtt 1416
```

```
Qy 1399 ggcgaagaaggctcaagcgtaacttgagcccaagaccataataatacaagt 1451
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1417 ggccttaactctttgagaactacacagatacaagacgggtgcactacgcct 1469
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-09-533-559-5519
; Sequence 5519, Application US/09533559
; GENERAL INFORMATION:
; APPLICANT: Randy M. Berka
; APPLICANT: Michael W. Rey
; APPLICANT: Jeffrey R. Shuster
; APPLICANT: Sakari Kauppinen
; APPLICANT: Ib Groth Clausen
; APPLICANT: Peter Bjarke Olsen
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 5849.200-US
; CURRENT APPLICATION NUMBER: US/09/533,559
; EARLIER FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 09/273,623
; NUMBER OF SEQ ID NOS: 7860
; SOFTWARE: FastSeq for Windows version 4.0
; SEQ ID NO 5519
; LENGTH: 2022
; TYPE: DNA
; ORGANISM: Aspergillus oryzae
US-09-533-559-5519

Query Match 16.8%; Score 245.8; DB 20; Length 2022;
Best Local Similarity 52.3%; Pred. No. 4e-57; Indels 36; Gaps 8;
Matches 750; Conservative 0; Mismatches 647;

Qy 34 ctcttcacacaaatgaattcgtctctcccaagggtccgagagatgaattaacgtcacgaac 93
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 451 ctgttcacacaaatgaattcgtctcccaagggtccgagagatgaattaacgtcacgaac 510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 94 ccgtgggaacaaatccacgttgccactgatgttcacgttagccaaacgcccgcgactgcac 153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 511 ccacacaacagagaagccgattg---tagccgtccagggagcgagcgaagaatgatcgaa 567
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 154 agtgcagtagcgtctcgttgacggcggtccaaagggcccccattggaagaagtccacagt 213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 568 gatgcgt---caaggcagcagcgcgccttcgaggggtgagtgagcaagtcactccc 624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 214 gcacacgcgcgcgtgcatgcttaagtgcggaacctgcgcgagagaagaacgcgcgagaag 273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 625 tcggaacgtggccgcctcctcgtccagctggccgacacctcttcgagcgtgtagagcagatc 684
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 274 ctgcgtcgtgagtcgctgcgcacccggttagaccggtgctgatcatcactcatttcgac 333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 685 ctggctgctattgaagcccttgacaaacgcaa---ggcctcaccaatggctaaaggccat 741
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 334 attccaaacatggtctcgtgttctcactactgacaggtggcgacaaagatcccgga 393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 742 gttgcgcgcgcgcgcgtcctcgtcctactactcgttggtggcggaagaattcacgcgc 801
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 394 aagacctcttcgaggaacacggaacgcgaattgg---cgttacgagccgattgggggtg 450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 802 cagaccttgacacaaacctgagtctcttactacacccgcgcagagccattggtgtc 861
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 451 tgtcgtgattccacgtggaacgcgactttcttactacgtcgtcgttggaagatagcccc 510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 862 tgtgtgcagatcccccgtggaactccctctctctctgatgtggtctctggaagattggaact 921
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 511 gccctgcgcgcgcgtcgtcctcctcactcctcctcctcctcctcctcctcctcctcctcct 570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 922 gccatcgtcccgtaacgtcgtgttccctcaagactgctgacagacccccctcgtcgtg 981
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 571 ctgggctcgtcctctcttcgcagaaacgcggaattccctcctcctcgtgagtgagttctc 630
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 982 ctctatgcgcgaagctcatcaaggaagcttccctcccgctgggtggttaatactctc 1041
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 631 acttgg---agcacgagtgacgggtgaagcatctggcgctgcacatcgacattgcgaagatc 687
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1042 tctggttcggcgtgtcgcgggtgctgtatctctagccacatggatcgacaagatt 1101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 688 agcttcacaagatctgtcggcggtggcgcgctcaagcaagcaacacacacacacacacac 747
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1102 gcttcaccggttcgacctggttgctgtagtattctccagcgccgccaagagacaac 1161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 748 atgaagcgctcactctagaaactgggggaaagccaaacacacacacacacacacacacac 807
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1162 ctgaagaagtgctcactctgagctgggtggcaagtccttaacattgtctcgcagacgc 1221
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 808 ctgcaacggcagtcgggggaatcgcaaa-----ggattctcaaaatctcggcaaat 861
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1222 gacatcgataacgctatctctgttccaaacttcggtatcttctcaacacacacacacac 1281
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 862 tgggtcccccctcctctgttctgtagtgcgaatgggggaaatttagcggagaaatctccatgga 921
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1282 tgcctggcggttcccgtaacctgtccagggagggcatccacgacaagctcgttgcgcgc 1341
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 922 gtcgctcatgctcatttgagagctgtcagatggtctggccagaacacacacacacacac 981
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1342 ttcaaggagcgcgcgtcgcaacaagc-----tcggcaacccctcctcactgc 1389
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 982 aagagacgcagatggtccctctcgtcgaagaagtcctcagacagagctcttgggttaacatt 1041
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1390 gacaccttcagggccctcaggtctcgcagctgctgttgcgcgcatcttgagctacatc 1449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1042 gacgtggcaagata---ccgcgcagctcctcactcctcctcctcctcctcctcctcctc 1098
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1450 aaccacggtgaagcagaagggtctaccgttgcacccggtggtagcgttcacggaaccgag 1509
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1099 ggattcgcgattgaacgcagcatattgttcaatcccaaacaccagggcagcaaaatttggttt 1158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1510 ggttactctcagcctacacgtcttccacgagtccacagcgacatgaagatcgccaag 1569
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1159 gagaagatcttggcccgctcttccattgaagcgttcaagacggaagaagagccatt 1218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1570 gaggagatctcggcccgctcgtccacatcccaagaattcaaggatgaggagggccatc 1629
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1219 gatttcacaaatgacacgacttatggctagcctcgttcatttataccacacacacacacac 1278
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1630 aagattggtaacgctcttcgttacggcttgcgcgcgcgtccacacacacacacacacac 1689
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1279 aggggtctcgtctcgttcgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1338
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1690 actgcacatccgctctccaaactctcctcgggtggaaacgcgtctggaatcacactgctacaac 1749
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1339 atccccgagacacaaactcgttttggcgcatgaaacaacatcggttcagggcagagagata 1398
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1750 atgataactaccaggtctcctcgttgggttcgaagagtcgggtcttggcggtgaactg 1809
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1399 ggcgaagaagggttcaggcgttacttggagcccaagacacacacacacacacacacacac 1451
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1810 ggcctcactcgccttgagaactacacacacacacacacacacacacacacacacacacac 1862
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 14
US-09-533-559-14
; Sequence 14, Application US/09533559
; GENERAL INFORMATION:
; APPLICANT: Randy M. Berka
; APPLICANT: Michael W. Rey
; APPLICANT: Jeffrey R. Shuster
; APPLICANT: Sakari Kauppinen
; APPLICANT: Ib Groth Clausen
; APPLICANT: Peter Bjarke Olsen
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 5849.200-US
```

```
; CURRENT APPLICATION NUMBER: US/09/533,559
; CURRENT FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 09/273,623
; EARLIER FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 7860
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 1710
; TYPE: DNA
; ORGANISM: Fusarium venenatum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1710)
; OTHER INFORMATION: n = A,T,C or G
US-09-533-559-14

Query Match          16.8%; Score 245.4; DB 20; Length 1710;
Best Local Similarity 53.48; Pred. No. 4.9e-57;
Matches 727; Conservative 0; Mismatches 600; Indels 34; Gaps 9;

Qy 112 gtgccaactgatgttcaacgtggaacacgcggccgatgtcgacagtcagtagccgcttcg 171
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 228 gtcatacctcgtctcgaggggtaccgaaaggatatacgacctggctgtcacccgtgcc 287
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 172 gtgcaggcgggtcaaaaaggcccatggaagaagtccacaggtgcacaacgcgcggctgc 231
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 288 cgcagggtttc---gacggcgagtggaagcagactactctcccaggctcggaggaacctc 344
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 232 atgcttaagtcgggaactcgcgcgagaagaacccgagaagctcgcgtcgtcgagctgc 291
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 345 ctctcaagcttgcgcagactcgcgcgagaagaacctcgcactcctcgcgcgtgttgagct 404
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 252 ctgcccacccgtagaccggtgtcatgatactactatttcgacattccaaacatgctctcc 351
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 405 ctgcacaacagggaagtcataccaa---cgtctggtggtgagtcggtgtgtgtgtgtggc 461
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 352 gtgttcctactactgacgggtggcgacaaagatcgcggaaagacaccttcccgaggagac 411
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 462 tgccttcgatactacgttggtggcgacgaagattgagggaagaccatcgatatgtct 521
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 412 aacggcaag---cgaattgctgtacgacccgatgggtgtgtgtgtgtgtgtgtgtgtgc 468
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 522 ccgagatgttccactacacccgatacgcgactatcggtctgctggtcagattatcccc 581
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 469 tggaaacgcagcttttcttactgctggtggaagatagccccgcctcgcgcgcggctgc 528
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 582 tggaaacttcccccttctctatgctggtggaagattggaactgctctggtgccactggtaac 641
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 529 tecttcattcaaaagcctcgggaataatccccgcctggg-cgattcgtggcctcgtctct 587
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 642 actatcgtcatgaaactgctgacagactcctctctcgccttcttcttcttctcactcagtt 701
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 588 ctctgcagaacccggatttctctctgagtgagtcgagttctca-----ctggagacaga 642
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 702 catcagagcgtggtgttctctcgcgtgtcttcaacctgttctcctcggntacngcaaga 761
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 643 gtgacggttgaagattggcgtgcacatgacattcgaagatcagcttcaacagatct 702
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 762 ctgnccggtgcgcctctctctcatatggacgttgcagaagatcgccttctcactcgtttcc 821
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 703 gtcggcggtggtgcgcgcgtcaagcaaacatcactcaagtcacacatgaagcgcgtcact 762
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 822 accgtcattggccgacagatcatgaaggctgctgtcttcttccacctcaagaaggtcacc 881
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 763 ctgaactcggggaaagcaaacatcgtcttcaacgaactctctcgaacgcagctgcg 822
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 882 ctgagcttggcgaagtcctcccaacatcgtcttctcgaagcagctgcacatcgaggagcc 941
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 823 ggggaatcggcaaa-----ggattctcaaaattcgggaaatttgggtcccccctctct 877
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 942 attaaactgggtcaactctcggtatctactacaaacacacgcccagtgctgtgtgcccgtacc 1001
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Qy 878 gtttgcgtagtcgaatggggaatttagcggagaaatttccattccatggagtcgctgctcat 937
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1002 cgtatcttctcagaggagtcacatctacg---acaagttctcgtcgtcgttcaagaagcga 1058
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 938 ttgagagctgtcagagatggtcttggccagaaacccatttggaaaccccaagagcagcagtc 997
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1059 gctgagagaaacaaggttggc-----gacctctcaacgaggaagacaccttccagggtc 1110
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 998 cttcgtcgcacaagctccacgtacgacagagctcttgggttaacattgacgttggcag---g 1054
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1111 cccaggtctctcagctccagtcacgacggtatcatgggtctacatcaagccggtgaaggacg 1170
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1055 ataccgcgcagctcctcactggtgttggtaagaaaggcgaacagggatttcgcgattgaac 1114
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1171 agggcgccactgttgagactggtgtgagcgtctcgtcgtgacaaaggttacttctcatcagc 1230
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1115 cgacgtatttgcataatcccaacccagcagcgaataatttgggttggagagatcttctggcc 1174
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1231 ccacattcttccaaagtcgcgcgcgcacatgaagatcatgcagaggaagagatttctcgcc 1290
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1175 ccgttcttgcattaaagcgttcaagacgggaagagccatttgcgatttgcgaatgaca 1234
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1291 ctgctcgtctatttccaaagtttaagacagcgtggtgaggtcttgccttgcctcagcagca 1350
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1235 cgacttatgctagcctcgtctcatttataccaaatctctcaacaggggtctcgtctct 1294
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1351 ccgttctcgtctcgtctcgtcttccacacaaagacctcaacacagccctccgcgcttt 1410
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1295 cgtcggcgtcgtgagaccggttggcgtctcgtatcgaacttcccccttcccccgagacacaa 1354
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1411 ccaacgctctcaagcggcgcagctctgggtgaactgctataacatgcttccatcagc 1470
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1355 ctccgttcttgcggcgtataaaacaatcggcctcagcagacagagctagcgaagaaggtcca 1414
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1471 ttccttggaggggtacaaagagagcggaaattggacgagagactcggggagggcagccctcg 1530
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1415 aggcgtacttggagcccaagaccattaatatccacgtcaac 1455
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1531 ccaactacacacanaacaaagctgtgtgcttccatcaagctgac 1571
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
PCT-US01-17253-89
; Sequence 89, Application PC/TUS01117253
; GENERAL INFORMATION:
; APPLICANT: Genaisance Pharmaceuticals
; APPLICANT: Duda, Amy
; APPLICANT: Finkel, Kevin
; APPLICANT: Kazemi, Amir
; APPLICANT: Messer, Chad
; APPLICANT: Sanchis, Angela
; TITLE OF INVENTION: Haplotypes of the ALDH5 Gene
; FILE REFERENCE: MMH-0719PCT ALDH5
; CURRENT APPLICATION NUMBER: PCT/US01/17253
; CURRENT FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/207,508
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 89
; LENGTH: 9589
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: allele
; LOCATION: (3589)
; OTHER INFORMATION: PS1: polymorphic base G or A
; NAME/KEY: allele
; LOCATION: (3593)
; OTHER INFORMATION: PS2: polymorphic base C or T
; NAME/KEY: allele
; LOCATION: (3645)
; OTHER INFORMATION: PS3: polymorphic base C or T
```

[illegible]

Db 5495 aatctggaayggaggagctgggtgaggatgggttaaggcctacacagaggtaaaga 5554
Qy 1436 ccattaatatccacgt 1451
Db 5555 cgtcacccatcaaggt 5570

Search completed: April 27, 2002, 01:33:51
Job time: 21995 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 00:19:28 ; Search time 150.7 seconds
(without alignments)
2386.247 Million cell updates/sec

Title: US-09-882-694A-4
Perfect score: 1464
Sequence: 1 atgggttttcgcctgcagcga.....ttccacgicacacatagatga 1464

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|----|--------------------------------------|
| 1 | 141.4 | 9.7 | 1842 | 4 | US-09-221-294-1 Sequence 1, Appli |
| 2 | 138.6 | 9.5 | 1455 | 4 | US-09-651-941-8 Sequence 8, Appli |
| 3 | 138.6 | 9.5 | 1524 | 4 | US-09-655-270A-8 Sequence 8, Appli |
| 4 | 138.6 | 9.5 | 12508 | 4 | US-09-655-270A-1 Sequence 1, Appli |
| 5 | 138.6 | 9.5 | 12523 | 4 | US-09-651-941-1 Sequence 1, Appli |
| 6 | 128.8 | 8.8 | 9515 | 1 | US-08-920-812-13 Sequence 13, Appli |
| 7 | 128.8 | 8.8 | 9515 | 1 | US-08-920-827-13 Sequence 13, Appli |
| 8 | 128.8 | 8.8 | 9515 | 1 | US-08-921-177-13 Sequence 13, Appli |
| 9 | 128.8 | 8.8 | 9515 | 1 | US-08-362-577C-13 Sequence 13, Appli |
| 10 | 128.8 | 8.8 | 9515 | 2 | US-08-920-828-13 Sequence 13, Appli |
| 11 | 121.6 | 8.3 | 4403765 | 4 | US-09-103-840A-2 Sequence 2, Appli |
| 12 | 121.6 | 8.3 | 4411529 | 4 | US-09-103-840A-1 Sequence 1, Appli |
| 13 | 115.6 | 7.9 | 1497 | 2 | US-08-513-841-4 Sequence 4, Appli |
| 14 | 115.6 | 7.9 | 1497 | 2 | US-08-696-834-4 Sequence 4, Appli |
| 15 | 115.6 | 7.9 | 1497 | 2 | US-08-942-673-4 Sequence 4, Appli |
| 16 | 115.6 | 7.9 | 1497 | 4 | US-09-118-317-4 Sequence 4, Appli |
| 17 | 115.6 | 7.9 | 4624 | 2 | US-08-696-834-5 Sequence 5, Appli |
| 18 | 86.4 | 5.9 | 728 | 4 | US-08-998-416-758 Sequence 758, App |
| 19 | 75.8 | 5.2 | 727 | 4 | US-08-998-416-738 Sequence 738, App |
| 20 | 63.8 | 4.4 | 1608 | 6 | 5221737-6 Patent No. 5221737 |
| 21 | 60.8 | 4.2 | 662 | 1 | US-08-322-742-16 Sequence 16, Appli |
| 22 | 58.2 | 4.0 | 2400 | 4 | US-08-930-001-1 Sequence 1, Appli |
| 23 | 56.2 | 3.8 | 402 | 1 | US-08-346-611-1 Sequence 1, Appli |
| 24 | 56.2 | 3.8 | 402 | 2 | US-08-794-494-1 Sequence 3, Appli |
| 25 | 55.6 | 3.8 | 402 | 1 | US-08-346-611-3 Sequence 3, Appli |
| 26 | 55.6 | 3.8 | 402 | 2 | US-08-794-494-3 Sequence 3, Appli |
| 27 | 53.8 | 3.7 | 595 | 1 | US-08-076-091C-5 Sequence 5, Appli |

| | | | | | |
|----|------|-----|-------|---|------------------------------------|
| 28 | 53.8 | 3.7 | 595 | 1 | US-08-285-641-5 Sequence 5, Appli |
| 29 | 37.2 | 2.5 | 68750 | 3 | US-09-335-409-1 Sequence 1, Appli |
| 30 | 37.2 | 2.5 | 68750 | 4 | US-09-568-102-1 Sequence 1, Appli |
| 31 | 37.2 | 2.5 | 68750 | 4 | US-09-567-969-1 Sequence 1, Appli |
| 32 | 37.2 | 2.5 | 68750 | 4 | US-09-568-480-1 Sequence 1, Appli |
| 33 | 37.2 | 2.5 | 68750 | 4 | US-09-568-486-1 Sequence 1, Appli |
| 34 | 37.2 | 2.5 | 68750 | 4 | US-09-568-472-1 Sequence 1, Appli |
| 35 | 37.2 | 2.5 | 71989 | 4 | US-09-443-501A-2 Sequence 2, Appli |
| 36 | 37.2 | 2.4 | 4931 | 4 | US-08-726-320-2 Sequence 2, Appli |
| 37 | 34.8 | 2.4 | 4931 | 4 | US-09-208-716-2 Sequence 2, Appli |
| 38 | 34.2 | 2.3 | 1680 | 4 | US-09-603-185-1 Sequence 1, Appli |
| 39 | 34 | 2.3 | 1715 | 4 | US-08-757-230A-1 Sequence 1, Appli |
| 40 | 34 | 2.3 | 1715 | 5 | PCT-US95-02315-1 Sequence 1, Appli |
| 41 | 34 | 2.3 | 3664 | 1 | US-07-880-913-1 Sequence 99, Appli |
| 42 | 33.6 | 2.3 | 613 | 4 | US-09-227-357-99 Sequence 5, Appli |
| 43 | 33.6 | 2.3 | 1251 | 2 | US-08-851-088-5 Sequence 1, Appli |
| 44 | 33.6 | 2.3 | 5535 | 1 | US-08-089-755A-1 Sequence 1, Appli |
| 45 | 33.6 | 2.3 | 5535 | 1 | US-08-089-755A-4 Sequence 4, Appli |

ALIGNMENTS

RESULT 1
US-09-221-294-1
; Sequence 1, Application US/09221294
; Patent No. 6266138
; GENERAL INFORMATION:
; APPLICANT: Riccardo Dalla-Pavera and
; APPLICANT: Alessandro Massimo Gianni
; TITLE OF INVENTION: A Retroviral Vector Capable of Transducing the
; TITLE OF INVENTION: Aldehyde Dehydrogenase-1 Gene and Uses of Said
; TITLE OF INVENTION: Vector
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM 330 466 DX2
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,294
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 42990-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1842 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: N
; ANTI-SENSE: N
; FRAGMENT TYPE: N-terminal
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1568
; US-09-221-294-1

Query Match 9.7%; Score 141.4; DB 4; Length 1842;

[illegible]

```
Db      1383   CCATAACAATCTCCTCTGCTCTCCAGGCAGGAACAGTGTGGTGTAATTGCTATGCCGTGG    1440
Qy      1343   ccgagacacaaaacctccgttttgaggcgcatgaacaatatcggcttcaggcacagagactaggcg    1402
          |||         |||||         |           |||||         |||         |||
Db      1443   TAGTGTCCCAGTCGCCCGCTTTGTGTGGATTCAAGATGCTCTGGAAATGGAAGAAGACTGGGAG    1502
          |||         |||||         |||||        |||         |||         |||
Qy      1403   agagaaggctcaaggcgctacttggagcccacaagaccatta    1441
          |||         |||||         |||||         |||||         |||         |||
Db      1503   AGTAGCGTTTTCCATGAATATACAGAGGTCAAAACAGTCA    1541
          |||         |||||         |||||         |||||         |||         |||

RESULT       2
US-09-651-941-8
; Sequence 8, Application US/09651941
; Patent No. 6355470
; GENERAL INFORMATION:
; APPLICANT: ROUVIER, PIERRE E
; APPLICANT: WALTERS, DANA M
; APPLICANT: RAINER, RUSS
; TITLE OF INVENTION: Genes Encoding Picric Acid Degradation
; FILE REFERENCE: BC1022 US NA
; CURRENT APPLICATION NUMBER: US/09/651,941
; CURRENT FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,545
; PRIOR FILING DATE: 1999-10-03
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Rhodococcus erythropolis HL pm-1
US-09-651-941-8
```

| Query Match | 9.58; | Score 138.6; | DB 4; | Length 1455; |
|-----------------------|-----------------|--|------------|--------------|
| Best Local Similarity | 48.18; | Pred. No. 6.6e-32; | | |
| Matches 686; | Conservative 0; | Mismatches 704; | Indels 35; | Gaps |
| Qy | 34 | ctctctcatcaacaatgaattcgtctctccaaagggtccgagagattaaagctcaacgaac | 93 | |
| Db | 28 | ctcgtcatcgcgcgaccaactgaacccatcgtctcgacggggcgaccttcgactcgatcaaac | 87 | |
| Qy | 94 | ccgtggagcaaatccacccgttgcaactgatgttcaactgggccaacgcgcgcgactgtcgac | 153 | |
| Db | 88 | ccggccgacgggtcgcaactcgacca--gcctccggaggccacggccgcgcgaactcgcg | 144 | |
| Qy | 154 | agtcagtagccgcttcggttcagggcgtcaaaaaggcccatgaagaaggttcaacaggt | 213 | |
| Db | 145 | cgtcgggtcgaagccgcgaaggcggcgcca-----ggacgtggcgacatgcgcgcgcg | 198 | |
| Qy | 214 | gcacaacgcggcggtgcatactctaagtctcggggaacctcgcggaagaagaacgcgcgagaag | 273 | |
| Db | 199 | gccacgcaaccccctcgtatcttcgcctacccgcgctgactcgagggaacacaagaccgag | 258 | |
| Qy | 274 | ctcgtcgtctcggagtcgctgccacccgttagaccgggtgctgactgatactcaatttcgac | 333 | |
| Db | 259 | ctgcgccagctgcagatcggagacatggcgaagcccatccgcgactcgtcgggatacgac | 318 | |
| Qy | 334 | attccaaacattggtctcgtgttcgtctactatgcaggctgggcgcgacaagaatgcgcgga | 393 | |
| Db | 319 | ctgcgatcatgactcgagacgtcgtcgactctcggcgcctcgtgaccaagatcgagggc | 378 | |
| Qy | 394 | aagacatttcccgaggcaacgcgaacgcgaatg--gcgttacgagcccatgggggtg | 450 | |
| Db | 379 | cgaacgacgcgcgcgcgcgcttctctcaactacacccctcgtgtagccgatactggttg | 438 | |
| Qy | 451 | tgtcgtgattgcagactggaacgcgacttttcttactcgttcggctggaagaatacccccc | 510 | |
| Db | 439 | gtggggcgcaacactccctggaaatttctcgtcagtcgagcggttcggaagatcgcccg | 498 | |
| Qy | 511 | gccctcgcgcgcgctcgtctctcaacttcaacagcctcggaataatccccctgggcgtt | 570 | |
| Db | 499 | gctcttgcatggcaaacgccatcgtgctgaagcctgcgcagcttcgcacccactcgtgcc | 558 | |

GENERAL INFORMATION:
APPLICANT: Ohno, Tsuneya
APPLICANT: Matsuhisa, Akio
APPLICANT: Uehara, Hirotsugu
APPLICANT: Eda, Saji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/920,812
FILING DATE: 29-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,577
FILING DATE: 27-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 9515 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Pseudomonas aeruginosa
STRAIN: Clinical isolate P2-2
US-08-920-812-13

Query Match 8.8%; Score 128.8; DB 1; Length 9515;
Best Local Similarity 48.2%; Pred. No. 1.4e-28;
Matches 641; Conservative 0; Mismatches 667; Indels 22; Gaps 9;

| | | | |
|----|-----|---|-----|
| Qy | 135 | caacgcgagcagatgcacagtcagtcgcttcggtgcagggcggtcaaaaaggccgc | 194 |
| Db | 641 | CGATCGCGCGACGTGACGCGCAGTGGCGCGCGCGCGCGCGCTTCGACGAAGGCC | 700 |
| Qy | 195 | atgaaagaattcacaggtgcacacgcgcgcgtgcatgcttaagtgcggacctcgc | 254 |
| Db | 701 | CTGGCGGGTCTCCCGCGTCGAGCGC-AAGCGCGTGTCTGCGCGTGGCGGAGCTGAT | 759 |
| Qy | 255 | cgagaagaacgcgcagtcgctgctgagtcgctgcacccgcgcggtagaccggtgctc | 314 |
| Db | 760 | GCTGGCCCATCGGAAGAGTGGCGTGTCTGACTGCTGTAACATGGGCAAGCCGTGAT | 819 |
| Qy | 315 | gatgatcactcatttcgacattccaaacatgctccgtgttttcctactatgaggtgc | 374 |
| Db | 820 | GGAGCGCTGGNAKATGATATACCCGGCGCGCCGCGCGCTTCCTGCTGATGCGGAAG | 879 |
| Qy | 375 | ggccgacaagatcgcgggaagaccttc---ccgaggacaacgcgcgaagccgaattg | 931 |
| Db | 880 | CCTCGAAGCTCTACGACAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 939 |
| Qy | 432 | ttacgagccgagtggt | 491 |

| | | | |
|----|------|--|------|
| Db | 940 | CCGCGTGCCTCGGGGGTATCGCGCGGTGTGCGGTGGAACTTCCCGCTCGACATGCC | 999 |
| Qy | 492 | cggctggaagatagccccccgcctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc | 551 |
| Db | 1000 | CGCTGGAAGCTCGCCCGCGCCCTGCGCGGCACTCGGTGGTGGTCTCAACGCGCGCA | 1059 |
| Qy | 552 | gaaatccccgcgtggtctctggtgctcctctctctctctctctctctctctctctctcc | 611 |
| Db | 1060 | GCATTCGCCGTTCTCGCCCTGCGCCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCG | 1119 |
| Qy | 612 | tggagtcgtgagtcctcctcactgagcagag---tgacgggtgagcagtcggtcgcga | 668 |
| Db | 1120 | AGGCGTCTCAAGCTGTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCTGCA | 1179 |
| Qy | 669 | catggacattggaagatcagcttcacagatctctcggcggtggtggtggtggtggtggt | 728 |
| Db | 1180 | CCGCGAGGTGACCGACTGGTGTTCACCGCTCCACCGAGTTCGCGAGTACTTCAATGCA | 1239 |
| Qy | 729 | agcaactcaagtcacacatgaagcgcgtcactctcactctcactctcactctcactctc | 788 |
| Db | 1240 | GTATTCGCGCAATCAAGCTCAAGCAGGTCTGGTGGAGTGGCGGTAAAGAGTCCGAA | 1299 |
| Qy | 789 | cgtcttcacagagctcctctcgaacgcgcagtcggggggaatcgcaaggatttctcaa | 848 |
| Db | 1300 | CCTGGTCTTCGCCGATTGCGCGCTTTCACCTGCGCGGGAAGGCGCTTCGGCAT | 1359 |
| Qy | 849 | attcgggcaaatgttggtccccccccctctgttctgtctgtctgtctgtctgtctgtct | 908 |
| Db | 1360 | TTTCTT-CAATCAGGCGGAGTGTGTTTCGCGCAACTCGCGCTTGTCTGG---TGGAGCCTT | 1415 |
| Qy | 909 | gaaatccatggagtcgctcatttggaggctgtcagagatggttggccagaa | 968 |
| Db | 1416 | CGATCCACGACGAGTGTGCGAGCGCTCTGTCGCAAGCGCGCGCACATGGCAGCGCGCG | 1475 |
| Qy | 969 | cccatgggaacccaag--aggacgatggtcctctctctctctctctctctctctctctct | 1026 |
| Db | 1476 | ATCGGTGGACCG | 1535 |
| Qy | 1027 | gtcttggttaacattga-cgttggcaaggatataccgcgcgcgcgcgcgcgcgcgcgcgc | 1085 |
| Db | 1536 | ATTCTCGCGCCATCGAGCG | 1595 |
| Qy | 1086 | aaaggcgcaaggggattcg-----cgattgaaccgcagcagatattgtcaatcccaac | 1138 |
| Db | 1596 | AGTTGACGATCAACGCTTCGGACAACTTCATCGAACCGCACCTGTTCCGGCAGCTACGCC | 1655 |
| Qy | 1139 | caggcgcaaaatttggtttgaggagatcttggccgcgcgcgcgcgcgcgcgcgcgcgcgc | 1198 |
| Db | 1656 | CGACATGACGTGCG | 1715 |
| Qy | 1199 | agcggaaagaaggcagtcattgagattgccaatgacacgacttatgggtgagtcggtca | 1258 |
| Db | 1716 | ACTCCGAGGACGAGGCGCATACGCTGGCCAGGACAGCGCTACGCGCTCGCGCGCTCGC | 1775 |
| Qy | 1259 | ttataccaaatctcacaagggtctcgtgtctgtctgtctgtctgtctgtctgtctgtctgt | 1318 |
| Db | 1776 | TGTGGAGCGACGACCTGCG | 1834 |
| Qy | 1319 | tctcgatcaacttccccctttatcccccgagacacaaactccgtttggcgagcatgaaacat | 1378 |
| Db | 1835 | TGTCGCTGAATACCGTGGACGCGCTGGAGCTGCGCGTGCCTTTCGCGCGCGCGCGCGAGT | 1894 |
| Qy | 1379 | cgggctcagcagagagctagcgaagaaggaaggaaggaaggaaggaaggaaggaaggaag | 1438 |
| Db | 1895 | CCGGCTTCGGTTCGCGACCTGTCGCTGCTTCCTTCGACAACTACACCGCATTTGAAGACGA | 1954 |
| Qy | 1439 | taatatcca | 1448 |
| Db | 1955 | CTTGGTTCCA | 1964 |

RESULT 7
US-08-920-827-13

```

; Sequence 13, Application US/08920827
; Patent No. 5770375
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotosugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,827
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9515 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
; STRAIN: Clinical isolate P2-2
; US-08-920-827-13

Query Match      8.8%; Score 128.8; DB 1; Length 9515;
Best Local Similarity 48.2%; Pred. No. 1.4e-28;
Matches 641; Conservative 0; Mismatches 667; Indels 22; Gaps

Qy 135 caacgcggccgagtcgacagtgtagcagccgttcggtgagcggtgcacaaaggsgccc 194
Db 641 CGATCGCCGCGACGCTGACGCGGCGAGCTGGCGCGCGCGCGCGCTTCGACGAAGGCGC 700
Qy 195 atggaagaagttcacaggtgcacacgcgcgcgtgcgtcttaagttcgcggaactcgc 254
Db 701 CTGGGCGCGTCTCGCCCCGGTCCAGCGC-AAGCGCGTCTCTGGCGCTGGCCGAGCTGAT 759
Qy 255 cgagaagaacccgagagtcgctcgtctgagtcgctccccaccggtagacgggtgc 314
Db 760 GCTGGCGCCATCGCGAAGAGATGGCGGTGCTCGACTCGCTGAACATGGGCAAGCGGTGAT 819
Qy 315 gatgatcaactatttcacattccaaacatgctccgtgtttcgtactactatcgagctg 374
Db 820 GGACGCTTGGAAACATCGATGTACCCCGCGCGCGCCACGCTCTTCGCTTGGTATCGGGAAG 879
Qy 375 ggccgacagaatcgcgggaagacctttc---ccgaggacacgcggaagccgaattggcg 431
Db 880 CTTGACAAAGCTCTACGACAGCTCGCGCGCGCGCGCGCGACGACCTGGCCACCATTAC 939

```


| Query Match | 8.8% | Score 128.8 | DB 1 | Length 9515 |
|-----------------------|---|-------------------|-----------|-------------|
| Best Local Similarity | 48.2% | Pred. No. 1.4e-28 | | |
| Matches 641 | Conservative 0 | Mismatches 667 | Indels 22 | Gaps |
| QY 135 | caacgcgcgcattgcacacagtcagtagccgcttcgggtgcaggcggtgcacaaagagccc | 194 | | |
| DB 641 | CGATGCGCGGACGTCCGACCGGACGTGGCGCCGCCGCCGCCCTTCGACGGAAGGCC | 700 | | |
| QY 195 | atggaagaagttccacaggtgcacaacgcgcgcggtgcagtgttaagttcgcggacctcgc | 254 | | |
| DB 701 | CYGGCGCGCTTCGCCCCGGTTCGAGCC-AAGCGGTGCTTCGCCCTGGCCGACGTGAT | 759 | | |
| QY 255 | cgagaagaacccgagaagctcgctcgctgtagtcgctgcccaccggtagacgcggtgc | 314 | | |
| DB 760 | GCTGGCCCATCGGAAGAGCTGGCGCTGCTCGACTCGCTGAACATGGCGAAGCGGTGAT | 819 | | |
| QY 315 | gatgatacactcatttcacattccaaacatgctccatgtttcgctactatcacagcgtg | 374 | | |
| DB 820 | GGACGCTTGGACATCATGATTACCGCGCGCCGCCACGCTCTTCGCGCTGGTATCGGGAAG | 879 | | |
| QY 375 | ggccgacaagatccgcggaagaccttc---ccgagacacgcggaagccgaattggcg | 431 | | |
| DB 880 | CCTCGACAAGCTCTACGACAGGTTCGGCGCGCCGCCACGACAGACCTTGGCCCAATTAC | 939 | | |

| | | | |
|--|------|--|------|
| Db | 1219 | ACCGTCGATGAGCGATCGCGATTGCCAATGACAGCGTTTACGGCTTGGCGCGATCGGTC | 1278 |
| Qy | 1261 | tataccaaatctctcaacagggttctcgtgtctctgtcgcgcgtcgcagaccggtggcgctc | 1320 |
| Db | 1279 | TGGAGCAAGGATATCGACAAGGCGCTTGCCTCGTGCACCGCTCGTGTTCGTGCGCGCGCTTC | 1338 |
| Qy | 1321 | tcgatcaactctccctttatctcccccgcgcagacaaactccgtttggcggcgatgaacaatcg | 1380 |
| Db | 1339 | TGGGTGAACACCATCATGAGCGGTGGTCCCGAGACGCCGCTGGGTGGTTTCAAGCAGTCG | 1398 |
| Qy | 1381 | ggctcaggcagagactagcgcaagaagggtctcaggcgtactctggagagcccaagaccatt | 1440 |
| Db | 1399 | GGCTGGGGCGGTGAGGCCGCTGTACGGCGCTTGAGGAATATACGCAGATCAAAATCTGTC | 1458 |
| Qy | 1441 | aatatcca 1448 | |
| Db | 1459 | CATATCGA 1466 | |
| RESULT 14 | | | |
| US-08-696-834-4 | | | |
| ; Sequence 4, Application US/08696834 | | | |
| ; Patent No. 5834263 | | | |
| ; GENERAL INFORMATION: | | | |
| ; APPLICANT: Niwa, Mineo | | | |
| ; APPLICANT: Saito, Yoshimasa | | | |
| ; APPLICANT: Ishii, Yoshinori | | | |
| ; APPLICANT: Yoshida, Masaru | | | |
| ; APPLICANT: Hayashi, Hiromi | | | |
| ; TITLE OF INVENTION: Method for Producing 2-Keto-L-Gulonic Acid | | | |
| ; NUMBER OF SEQUENCES: 48 | | | |
| ; CORRESPONDENCE ADDRESS: | | | |
| ; ADDRESSEE: Obion, Spivak, McClelland, Maier & Neustadt, | | | |
| ; ADDRESSEE: P.C. 1755 Jefferson Davis Highway, Suite 400 | | | |
| ; STREET: 1755 Jefferson Davis Highway, Suite 400 | | | |
| ; CITY: Arlington | | | |
| ; STATE: Virginia | | | |
| ; COUNTRY: USA | | | |
| ; ZIP: 22202 | | | |
| ; COMPUTER READABLE FORM: | | | |
| ; MEDIUM TYPE: Diskette - 3.50 inch, 1.44 Mb storage | | | |
| ; COMPUTER: IBM PC compatible | | | |
| ; OPERATING SYSTEM: PC-DOS/MS-DOS | | | |
| ; SOFTWARE: | | | |
| ; CURRENT APPLICATION DATA: | | | |
| ; APPLICATION NUMBER: US/08/696.834 | | | |
| ; FILING DATE: 24-SEP-1996 | | | |
| ; CLASSIFICATION: 435 | | | |
| ; PRIOR APPLICATION NUMBER: JP 28612/1994 | | | |
| ; FILING DATE: 25-FEB-1994 | | | |
| ; ATTORNEY/AGENT INFORMATION: | | | |
| ; NAME: | | | |
| ; REGISTRATION NUMBER: | | | |
| ; REFERENCE/DOCKET NUMBER: | | | |
| ; TELECOMMUNICATION INFORMATION: | | | |
| ; TELEPHONE: (703) 413-3000 | | | |
| ; TELEFAX: (703) 413-2220 | | | |
| ; TELEX: 248855 OPAT UR | | | |
| ; INFORMATION FOR SEQ ID NO: 4: | | | |
| ; SEQUENCE CHARACTERISTICS: | | | |
| ; LENGTH: 1497 base pairs | | | |
| ; TYPE: nucleic acid | | | |
| ; STRANDEDNESS: double | | | |
| ; TOPOLOGY: linear | | | |
| ; MOLECULE TYPE: genomic DNA | | | |
| ; ORIGINAL SOURCE: | | | |
| ; ORGANISM: Gluconobacter oxydans | | | |
| ; STRAIN: T-100 | | | |
| ; FEATURE: | | | |
| ; NAME/KEY: CDS | | | |
| ; LOCATION: 1..1494 | | | |

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 00:40:26 ; Search time 706.01 seconds
(without alignments)
3560.237 Million cell updates/sec

Title: US-09-882-694A-4
Perfect score: 1464
Sequence: 1 atggttttttcgctgacga.....ttccacgtcaacatagatga 1464

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------|
| 1 | 1464 | 100.0 | 1464 | 22 | Exophiala spinifera |
| 2 | 267.6 | 18.3 | 1491 | 16 | Cladosporium herba |
| 3 | 245.8 | 16.8 | 2022 | 21 | Aspergillus oryzae |
| 4 | 245.4 | 16.8 | 1710 | 21 | Fusarium venenatum |
| 5 | 239.2 | 16.3 | 1488 | 16 | Alternaria alternata |
| 6 | 217 | 14.8 | 1980 | 23 | Drosophila melanog |
| 7 | 211.6 | 14.5 | 3442 | 23 | DNA encoding novel |
| 8 | 194.8 | 13.3 | 2700 | 7 | Sequence Constitut |
| 9 | 191 | 13.0 | 3463 | 21 | Human cancer assoc |

| | | | | | | |
|----|-------|------|---------|----|----------|--------------------|
| 10 | 190 | 13.0 | 1646 | 21 | AAF15973 | Human prostate can |
| 11 | 184.6 | 12.6 | 2036 | 20 | AAZ33590 | Human breast tumou |
| 12 | 182 | 12.4 | 1541 | 22 | AAF82082 | Human aldehyde deh |
| 13 | 179.6 | 12.3 | 2071 | 21 | AAF18274 | Lung cancer associ |
| 14 | 174.2 | 11.9 | 1772 | 21 | AAZ56975 | Arabidopsis aldehy |
| 15 | 172.2 | 11.8 | 1863 | 21 | AAZ56973 | Arabidopsis aldehy |
| 16 | 172.2 | 11.8 | 2182 | 23 | ABL06647 | Drosophila melanog |
| 17 | 168.6 | 11.5 | 109519 | 22 | AA508693 | Micromonospora DNA |
| 18 | 161.6 | 11.0 | 6291 | 23 | ABL06646 | Drosophila melanog |
| 19 | 156.6 | 10.7 | 6153 | 23 | ABL10878 | Drosophila melanog |
| 20 | 154 | 10.5 | 1614 | 21 | AAZ56974 | Arabidopsis aldehy |
| 21 | 153.6 | 10.5 | 1589 | 21 | AAZ44854 | Arabidopsis thalia |
| 22 | 141.4 | 9.7 | 1842 | 15 | AAQ72450 | Aldehyde-dehydroge |
| 23 | 138.6 | 9.5 | 1524 | 21 | AA53945 | ORF4 sequence enco |
| 24 | 138.6 | 9.5 | 12508 | 15 | AA53941 | 12.5 kb picric aci |
| 25 | 128.8 | 8.8 | 9515 | 15 | AAQ55145 | Pseudomonas aerugi |
| 26 | 126 | 8.6 | 1473 | 23 | AA554367 | Pseudomonas aerugi |
| 27 | 121.6 | 8.3 | 4403765 | 22 | AA199683 | Mycobacterium tube |
| 28 | 121.6 | 8.3 | 4411529 | 22 | AA199682 | Mycobacterium tube |
| 29 | 118.2 | 8.1 | 1473 | 23 | AA552323 | E. coli DNA for ce |
| 30 | 117 | 8.0 | 1446 | 22 | AAF61032 | P. putida KT2440-a |
| 31 | 117 | 8.0 | 1731 | 21 | AA50961 | Arabidopsis thalia |
| 32 | 117 | 8.0 | 1735 | 21 | AA44647 | Arabidopsis thalia |
| 33 | 115.6 | 7.9 | 1497 | 15 | AAQ74471 | L-sorbose dehydr |
| 34 | 115.6 | 7.9 | 1497 | 16 | AAQ04279 | Gluconobacter oxyd |
| 35 | 115.4 | 7.9 | 2735 | 23 | ABL03405 | Drosophila melanog |
| 36 | 114 | 7.8 | 3278 | 23 | ABL24684 | Drosophila melanog |
| 37 | 113.8 | 7.8 | 1425 | 22 | AAH22876 | R. ruber cdd gene |
| 38 | 113.8 | 7.8 | 10480 | 22 | AAH22877 | R. ruber 10kb gene |
| 39 | 110.8 | 7.6 | 4624 | 16 | AAQ04318 | Expression vector |
| 40 | 107.6 | 7.3 | 1226 | 21 | AAQ3109 | Arabidopsis thalia |
| 41 | 106.2 | 7.3 | 1232 | 21 | AA50967 | Arabidopsis thalia |
| 42 | 105.6 | 7.2 | 715 | 22 | AA541088 | cDNA encoding nove |
| 43 | 104 | 7.1 | 1452 | 23 | AA551421 | Pseudomonas aerugi |
| 44 | 100.6 | 6.9 | 1827 | 21 | AAQ37522 | Arabidopsis thalia |
| 45 | 97.8 | 6.7 | 1446 | 23 | AA556094 | Salmonella typhi D |

ALIGNMENTS

RESULT 1

AAQ02692
ID AAQ02692 standard; DNA; 1464 BP.

AC AAQ02692;

DT 02-MAY-2001 (first entry)

DE Exophiala spinifera aldehyde dehydrogenase coding sequence.

KW Fumonisin; degradative enzyme; transport enzyme; flavin monooxygenase;
aldehyde dehydrogenase; permease; p-glycoprotein; fumonisin esterase;
detoxification; mycotoxin; animal feed; human feed; silage;
transgenic plant; transgenic animal; microbial spray; ds.

XX Exophiala spinifera.

OS Key Location/Qualifiers

FT CDS 1..1464

FT /tag= a

FT /product= "E. spinifera aldehyde dehydrogenase"

XX WO200105980-A1.

XX 25-JAN-2001.

XX 14-JUL-1999; 99WO-US15824.

XX 14-JUL-1999; 99WO-US15824.

XX (PION-) PIONEER HT-BRED INT INC.

XX (CURA-) CURAGEN CORP.

AC AA086278;
XX
XX 25-OCT-1995 (first entry)
XX
XX Cladosporium herbarum allergen Clah53 cDNA.
XX
XX Fungal spore; allergen; Clah53; allergy; aldehyde dehydrogenase; ds.
XX
XX Cladosporium herbarum.
OS
XX
XX Key Location/Qualifiers
FH 1..1491
FT /*tag= a
FT /product= Clah53
XX
XX W09506121-A.
PN
XX
XX 02-MAR-1995.
XX
XX 24-AUG-1994; 94WO-AI00120.
XX
XX 27-AUG-1993; 93AT-0001725.
PR
XX (BIOM-) BIOMAY PROD N & HANDELSGES MBH.
PA
XX Achatz G, Breitenbach M, Ebner C, Hirschwehr R;
PI Kraft D, Lechenauer E, Oberkofler H, Prillinger H;
PI Simon B, Unger A;
XX
XX WPI: 1995-106850/14.
DR P-PSDB; AAR71891.
XX
XX Allergens derived from Cladosporium herbarum spores - also
PT recombinant DNA for expressing the allergens, useful for in vitro
PT allergy detection
XX
XX Claim 1; Page 8-9; 35pp; German.
XX
XX Spores of Cladosporium herbarum are the most common fungal spores
CC found in the air; they can cause allergic reactions. Various Clah
CC allergens and sequences encoding them have now been isolated. The
CC mature Clah53 allergen has mol. wt. 53 kD and is encoded by cDNA
CC sequence AA086278. The allergen has homology to aldehyde
CC dehydrogenases. Potential epitopic subfragments were identified by
CC computer analysis of the amino acid sequence. See AAR71892-R71906 for
CC potential B-cell epitopes and AAR72615-R72627 for potential T-cell
CC epitopes.
XX
XX Sequence 1491 BP; 325 A; 474 C; 395 G; 297 T; 0 other;
SQ

Query Match 18.3%; Score 267.6; DB 16; Length 1491;
Best Local Similarity 53.0%; Pred. No. 5.3e-70;
Matches 766; Conservative 0; Mismatches 654; Indels 26; Gaps 8;

QY 17 acgaatacaagagtgaactttcatcaacaatgaattcgtctctcctcaaggggtccgaga 76
DB 41 acgagcagccgacggcgtctcttcaacaacacgagttcgtcaagggccaaagagcaaga 100
QY 77 gattacacgtcaacaacccggggaacaaatccaccgttgccactgatgttcacgtggcca 136
DB 101 ccttcgattgatacaacccctccgacga---gagcgtgataccacaggtccacagggcca 157
QY 137 acgcccgcagtgatgcagagtgcagtagccgttcgttcgagggcggtcaaaaaggcccat 196
DB 158 ccgaagaagatgacacatccgctccgcccgcgcgcacagcctc---gagggtccat 214
QY 197 ggaagaacttcacaggtgcacaaacgcggcggtgcatgcttaagtctcgagacactgcgcg 256
DB 215 ggagactggagacacccgagaaacgtggcaagctgtcacaacacctgcgcaacctgttgcg 274
QY 257 agaaagaacgcgagaaagctcgtctgtgagtcgctgccaccggtagaccgggtgcga 316
XX

DB 275 agaagaacactgacctccttctgctgcggttgagtcgctcgacaacagcgcaagccacttcca 334
QY 317 tgatacctcatttcgacattccaaacatggtctcctggttctcgtactatcgaggtcggg 376
DB 335 tg---gcaagggtgacatcagatgtgcgtccgggtcctcagatactacggttggtggg 391
QY 377 ccgacaagatccgcaagaacacctttcccgaggacaacgcaagccgaattggcgttac- 435
DB 392 cggacaagatcaccgggcaagggtcaccactacgcacacccgcacacttccaactaagtcaga 451
QY 436 --gagccgagtggtgtgtgctggtattgccagctggaacggagacttttttaagtcog 493
DB 452 agggaccatgggtgttgccggtcagatcattccctggaacttccctctctatgctggg 511
QY 494 gctggaagatagccccgcctccgcgcgtcgtctccttcaatctcaagacctcgaga 553
DB 512 catggaagatcggcccggccattgtctgcggtcaacactgtcgtctgtgaagactgtcgagc 571
QY 554 aatccccgctggcggttctgggctcgtcctctcttcgcagaagccggagttccctctc 613
DB 572 agacctcttgggtggtctcgtcgtccagcctcgtcaagagcgcggtttccctcctg 631
QY 614 gagtgtgcagttcctcaactgagacaga---gtgacgggtgaagcattggcgtcgaca 670
DB 632 gtgtcatcaacgtcatttccgggttcggcaagggtcgtggtgcgctctctctctcaca 691
QY 671 tggacattgcgaagatcagcttcacaagatctgtcggcggtgcccgcgcgtccaagcaag 730
DB 692 tggacgtcgacaaggtggtcctccacgttccacgttgcgcgcgcacacatccctcaag 751
QY 731 caaactcaagtccaacatgaagcgcgtcactctagaactgagggggaagcccaacctcg 790
DB 752 ctgtgctctcttccaaacttgaagaggttccacctcggagtcggtggcagctcaccaca 811
QY 791 tcttcaacgaagctcctctcgaacgcgagtcgggggaatcggcgaagatttctcacaat 850
DB 812 ttgtcttcgagcgccgcatattgacaacgcatctcattggtcaacttcggtatctct 871
QY 851 tggggcaaatitgggtcccccctcctgttctgtagtgcaatgggggaaatttcgagaga 910
DB 872 tcaacacgcgcagtgctgctgtggttgcg-----gtgtgtacttcagagagagc 924
QY 911 aattccatggatcgtcgtcatggtcatttggag-gctgtcagagatggcttggccagaac 969
DB 925 atctacgacagttcgtccagaagtcaagagcgcgcacagaagaacgttgttggcgac 984
QY 970 ccattggaaaccaagagagcagcatggtcctctcgtcgaacaagtcctcgaacagagtc 1029
DB 985 ccttcgcgcgcgacacattccagggtcctcaggttcttccaaaggttcagtcgacgcgcat 1044
QY 1030 ttgggttaacattgacgttggcaagagatacc---gscgagctcctcactggcggttgtaga 1086
DB 1045 atggagatcatccagccgcgcaagagcgcggtgcccaccgtcgagagcgggtggagccgt 1104
QY 1087 aaggcgcaacagggattcgcgattgaaccgcgatatattgtcattcccaaacacagcgagc 1146
DB 1105 aagggtgacaagggtacttactttagcggccaccattctctccacgtccacagagagcatg 1164
QY 1147 aaaaattgggttggagagatctttggccccgctctgttcatttaagacgtctcaagacgaa 1206
DB 1165 aagatcgtgaagaagagatcttcggccccgtctgctgcatcgccaagttcaagaccaag 1224
QY 1207 gaagaggccattgagattgccaatgacacgacttatgggtcgtagctcgtgatttatacc 1266
DB 1225 gaggcgccatcaagctcggcaacgccagacacatcaggtctcgcgcgcgcgcgtccacacc 1284
QY 1267 aaatctcacaacaggggtctcgtgctcgtcgcgcgtcgcgagcgggtggcgctcgtgatc 1326
DB 1285 aagaacctcaacacgcgcagcatcgaggtctccaaagcgtctcgaagggcgccacgtctgggtc 1344
QY 1327 aacttccctttatccccgagacacaaactccgtttggcggtcatgaacaatcgggttca 1386
DB 1345 aacacttaacacacccctccaccaccagatgcggttcggtggtctacagaggtccggtatt 1404


```

Db 1390 gacacattcaggccctcagggtctcgcagctgcaggtttgacgcgcatcatgagtagcatc 1449
Qy 1042 gacgttggaagata---ccgcgcagctccctcactggcggttggtagaaaggcgacaag 1098
Db 1450 aaccaggttaagcagaaggtgtactacgttgcacccggtggtagcgtcaggaaccgag 1509
Qy 1099 ggattcgcgttgaacgcagcatatttgcataatcccaaacccgagcagcaaaatttggttt 1158
Db 1510 ggttacttcacagctccctcaccggttccacgagatgtccacagcagcatgaagatgccaa 1569
Qy 1159 gaggagatttggcccgcttcttcattaaagattcaagcagcaggaagagccatt 1218
Db 1570 gaggagatttggcccgcttccacacccagagattcaagaggttcaagatgaggaggccatc 1629
Qy 1219 gagattgccaatgacagcactatgggttagctagctcggtcatttataccaaatctctaac 1278
Db 1630 aagattggttaacagctcttcgtacggtcttgcgcgcgcgtccacacaaagacgttaac 1689
Qy 1279 aggggtctcgtgtctcgtcgcgctcgagaccggtggcggtctcgatcaactcccttt 1338
Db 1690 actgccatcgcgtctccaaactctccgggtcggaacctgtgataactgctacaac 1749
Qy 1339 atcccagagacacaaactcgttggcgcatgaacaaatcgggtcaggcagagagcta 1398
Db 1750 atgatacactaccaggtccctcgttgggttccagggatcggttggcgtgaactg 1809
Qy 1399 ggcgaagaaggctcaaggcgtacttggagcccaagaccatcaatccact 1451
Db 1810 ggtctacaccccttgagaactacacccagggtcaagactgtgactacccgct 1862

RESULT 4
ID AAF07491
XX AC AAF07491;
XX DE
XX DT 13-MAR-2001 (first entry)
XX KW Fusarium venenatum EST SEQ ID NO:14.
XX PT Multiple gene expression; filamentous fungal cell; EST;
XX KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
XX KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
XX KW culture condition; environmental stress; spore morphogenesis;
XX KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX OS Fusarium venenatum.
XX PN WO200056762-A2.
XX PD 28-SEP-2000.
XX PF 22-MAR-2000; 2000WO-US07781.
XX PR 22-MAR-1999; 99US-0273623.
XX PA (NOVO ) NOVO NORDISK BIOTECH INC.
XX PA (NOVO ) NOVO NORDISK AS.
XX PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX WP: 2000-594572/56.
XX PT Monitoring differential expression of genes in filamentous fungal cells
XX PT uses fluorescence-labeled nucleic acids isolated from the cells and a
XX PT substrate of expressed sequence tags -
XX Claim 86; Page 368; 3161pp; English.
XX The present invention describes a method for monitoring differential
XX expression of genes in a first filamentous fungal (FF) cell relative to

```

```

CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11853 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
XX SQ Sequence 1710 BP; 395 A; 508 C; 401 G; 399 T; 7 other;

```

```

Query Match 16.8%; Score 245.4; DB 21; Length 1710;
Best Local Similarity 53.4%; Pred. No. 2.8e-63;
Matches 727; Conservative 0; Mismatches 600; Indels 34; Gaps 9;

```

```

Qy 112 gtgcgaactgatttcaactggtgccaacgcgcgagtgatgcagagtgacatgcccttgc 171
Db 228 gtcatcaccctcgtctgcgagggtaccgaaaggaatgcagcctggctcaccgcctgc 287
Qy 172 gtgcagcggttcaaaaaggcccatggaagaagttcacaggttgcaacaacgcgcgcgc 231
Db 288 cgcaaggcttcc---gatggcgagtggaagcagactactccccccaggtcgaggcaacctc 344
Qy 232 atgcttaagttcgcgcgacctcgcgcgagaagaacgcgcgagactcgtcgtgtagctgc 291
Db 345 ctctcaagcttgcgcgacctcgcgcgagaagaacactcgcctcctcgcctgtagctgc 404
Qy 292 ctgcaccacggtagaccggtgtcgtatgcactcatttcgacattcccaaacatggctccc 351
Db 405 ctgcacaacgaaagtccatcaccaa---cgctcggggcgatgctcgtgtgtgtgc 461
Qy 352 gtgtttcgtactatgcagctggccgcacaagatgcgcgaaagacactttcccgagac 411
Db 462 tgccttcgatactacggtggttggccgacagatgaggggcaagaccatcgatatgtct 521
Qy 412 aacggcgaag---ccgaattggcgttacgagccgagtggtgtgtgtgtgtgtgtgc 468
Db 522 ccgagatgttccactacacccgacatccgagcctatcggtgtgtgtgtgtgtgtgtgt 581
Qy 469 tggaaacgcgactttttttagctgctggaagatagcccccgcctcaccgcgcgtgc 528
Db 582 tggaaactccctctcctcgtcgtggcattggaagatggacctcgtcgtcgcacgtgaac 641
Qy 529 tcttcattcttcaagcctcggagaaatcccccgctggg---cgttcgtggcctcgtcct 587
Db 642 actatcgtcatgaaactcgtcgcagactcctcttctcgtcgtcgtcgtcgtcgtcgt 701
Qy 588 ctgcgagagccgagattcctcctcgtgagtcgtagtctctca---ctggagacaga 642
Db 702 catcgacaggtgttctcctcgcggtgtcttcaacctgtcttccggttccngcaaga 761
Qy 643 gtgcggtgtaagcattggcgtcgcacatggaatgcgaagatcagcttcaagaatct 702
Db 762 ctgncgcgtgcgcctcctcctcactcagatgacgttgacaagaatcgctcactggttc 821
Qy 703 gtgcggtggtgcgcgcgcgttaagcaagcaactcaactcaagtcacatgaagcgcgtcact 762
Db 822 accgtcattggccgacagatcatgaaggtcgtgcttcttccacacctcaagaaggtcac 881

```


Db 576 acctctccgcatacattgctgcaagctgaccagagcggtttccaccagggtgt 635
Qy 618 cgtgcagttctcactcagcagcagtgga--cggtgaagcattgctgcacatgga 674
Db 636 catcaagctcactcgttttcgaaagatcgccggtgctgctccatgcacatgga 695
Qy 675 cattgcgaagatcagcttcaagatctgtcggcggtgctgcgcgcgaagaacac 734
Db 696 cattgacaagattcgtcttactgttcaaccgtgtcgccgctcaaatcatgaagtctgc 755
Qy 735 actcaagtccaacatgaagcgcgtcactctagaaactgggggaaagccaacca-----t 788
Db 756 ggctgctccaactgaagaagtcactcttgagctcgaggaagagcccaacaatgt 815
Qy 789 gctctcaacgaactcctctcgaacgcgagtcgagggaatcggaagattctcaca 848
Db 816 ctgcgcgaacgagattgacgaggtctatccactgggtcaactttggtatttactca 875
Qy 849 attcgggcaaatgtgggtcccccctcctgtttgttagtgcgaatgggaaatttagcgga 908
Db 876 ccacgacagcgtgtgtgctgttcgtctatctacgtccaagaagatctacgaca 935
Qy 909 gaaattccatggagtc--cgtcatgctcatttggaggctgtcagagatggcttggcag 966
Db 936 gttcatcagcgttcaagagcggctgctcagaacgctgtgttgacccattcgcgc 995
Qy 967 aaccatttgaaccaagagacgctgttccctctcgtgcacaagtccacgtacgcaga 1026
Db 996 gacac-----tccagggtccatcaagtctcgcagctcagttcgacgt 1038
Qy 1027 gcttggttaacatgacgttggcaagga---taccgcgacgtctcactcgtggtgt 1083
Db 1039 atcatgggtcactacgcagaggggcaagaagctctggcgacacatcgagactgttgcaac 1098
Qy 1084 agaagggcgacaaggtatcgogattgaaccgcagcatatttgcattcccaaacagagc 1143
Db 1099 cgtaaggtgacaaggttacttctacgagcccaaatcttcccaacgttaaccagagac 1158
Qy 1144 agcaaaattgttgtaggagatcttggcccgtcttgccttgcattgaagcgttcaagacg 1203
Db 1159 atgagattcagaagaagagatcttcggcccgtctgcacaaatctccaaagtccaagaca 1218
Qy 1204 gaagaagagccttgattgccaatgacacgactattggctagcctcgtctattat 1263
Db 1219 aagggcagcgtcatcaagattggcaacaacacacacatacaggtcttcggccgctgatac 1278
Qy 1264 accaaattctcaacaggggtctcgtgtctcgtcgtgcgtcgcagaccggtggtctcgt 1323
Db 1279 acatccaaactccaccactgcctcgaatgtgccaacgcgtcgcgtgagaaactgctgtg 1338
Qy 1324 atcaacttcccctttatcccgcgagacacaactccgttttgccgcatgaaacaatcgggc 1383
Db 1339 gtcaactctacaactcttcaactggtcagcttccccttcggagggtacaaggagctggt 1398
Qy 1384 taaggcagagactagcgaagaagggtcgaaggcgtacttggagccgaagaccattaat 1443
Db 1399 attgggcgcagtttggagagggcgcgctgtggaactacatcatcgaagaccgaacgctgt 1458
Qy 1444 atcc 1447
Db 1459 atcc 1462

RESULT 6

ABL10879
ID ABL10879 standard; cDNA; 1980 BP.
XX
AC ABL10879;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 27119.
XX

KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
DR P-PSDB; ABB66776.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 27119; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1980 BP; 471 A; 527 C; 540 G; 442 T; 0 other;

Query Match 14.8%; Score 217; DB 23; Length 1980;
Best Local Similarity 50.0%; Pred. No. 1.1e-54;
Matches 712; Conservative 0; Mismatches 690; Indels 21; Gaps 6;

Qy 32 aactcttcatacaaatgaattcgtctctccaaaggggtccgagagattaaagctcacga 91
Db 557 aacttttcatacaaatgagttcgttgattcgtgctgggcaagaccttgcaccttca 616
Qy 92 accgctgggacgaatccaccgttgcacactgattcactggccacgcggccgagtgcg 151
Db 617 atcc--ggctacatccaaggagattgtccaagtctccgagggagataaggctgatacg 673
Qy 152 acagtgagtagcgttccgtgcagggcgtcaaaaaggcc--atgagaagaagtcca 208
Db 674 acctggtgtgaagcggccaaagaagccttccatgcgactcgaatggcgcaaatlga 733
Qy 209 cagggtgcacacgcgcgcgcgtgcattgtcttaagtctcgacacctcgcgagagaagacgcgc 268
Db 734 gtctctgcagcgacccaattgatgaacaacttgcctcctgatggatcgggacaagg 793
Qy 269 agaagctcgtctctgagtagctgcccaacgcgtgagaccggtgctgatgatactcatt 328
Db 794 cttctcggccagcttgagagaccaggaacattggaagccctacactgag---gccctct 850
Qy 329 tcgacattccaacatggtctccgtgtttcgtactactatcaggtcggccgcgacaagatcg 388
Db 851 tcgactgtactactgactcgtgacgtgcagtactacgcgcgcgtggaccgacaagtct 910
Qy 389 ccggaaagacatttcccgagacaacgcgcaagccgaattggcgttaccagcgcgagggg 448
Db 911 tcggcgacacacttcccgctggaggcttcttccatgatcccccycaagcgcgctggcg 970

Query Match 13.3%; Score 194.8; DB 7; Length 2700;
Best Local Similarity 52.2%; Pred. No. 6.1e-48;
Matches 639; Conservative 0; Mismatches 548; Indels 36; Gaps 8;

QY 37 ttcatacaataaattcgtctctccaaaggggtccagagagataaacgtccacgaaccog 96
DB tttatcaaacacagtggttgaaagggcgttgagggcaagacgtccaggtccatcaacccc 953
QY 97 tggagcaatccaccgttgccactgatgttcacatgacacacgcccagatgcacagt 156
DB ttt 1007
QY 954 tcaaacga---gaagggtacactccgtcccaagggccaccaggaaggtgttga---t 1007
QY 157 gcagtagccgtctcgttgagggcgttcataaaagggccctatggagaagtccacagtgca 216
DB ttt 1007
QY 1008 gtgcgctgctgctgcctgctgctgttgaggggcccaggtgcacccctct 1067
QY 217 caacgcgcgcgtgcatgcttaatttcgagacctcgccagagaacccgagagagctc 276
DB ttt 1007
QY 1068 gacgttggaattttgtatcaaacagcgtggtgatctgatggagcgtgatctgacacctc 1127
QY 277 gctcgttgagtcgtgcccacccggtagaccggtgctgatgatcatcactatcgcacatt 336
DB gcgctatcgagtcctcgacaacggcgaaggctttcacatggcca---aggtcgatctt 1184
QY 337 ccaaacatgtctccgtgttctgactactatgcagcgtggtccgacaaagatcgccgaag 396
DB ttt 1007
QY 1185 gccaaactccattggtgtgctgatactacgctggctggcgacagaatccacggtccag 1244
QY 397 acctttcccgagacacacggcaagc---cgaattggtgttacagacgcatggggtgtgt 453
DB ttt 1007
QY 1245 accattgacacacaccccgagactttaccatacccccacacgagccggtgtgtgttc 1304
QY 454 gctggtattgcagctggaacgcgaactttctttacgtgcgtggaagatagccccgc 513
DB ttt 1007
QY 1305 ggtcagatctccctggaaacttcccctctctgatgtggtccctggagaatctggaccgct 1364
QY 514 ctgcgcgcgcgtgctctctcatcttcaaaagcctcgagaaatcccccgtggcgtgtctg 573
DB ttt 1007
QY 1365 gttgcgctgttaaacactgtgtcccaagacccgacagacccctctctcgcctt 1424
QY 574 ggcctgcctctctctcagaaagcggattccctccctcctgagctgctgagttccact 633
DB ttt 1007
QY 1425 tacgctgctgaagtgatacaagggtggtccattcccccgcgtggtgtgatcaacgtcactct 1484
QY 634 gg---agcacgagtgacgggtgaagcattggcgtcgcacatgacacattgcgaagatcag 690
DB ttt 1007
QY 1485 ggccttgccgtaccgctggtgtgctgccatccagccacatggacattgacaaaggtgcc 1544
QY 691 ttcacaagatcgtcggcgtggtccgcgcgtcgaagcaaacacacacacacacacacac 750
DB ttt 1007
QY 1545 ttcactggcttactctgttggaactacacacacacacacacacacacacacacacac 1604
QY 751 aagcgcgtcactctagaaactgggggaaagcccaacacacacacacacacacacacacac 810
DB ttt 1007
QY 1605 aagaaggtcactcttgagctcgttgcaagctcctcccaacacacacacacacacacacac 1664
QY 811 gaacggcagctcgggggaatcggcaaaagat-----ttctcaaaattcgggcaaaattgg 864
DB ttt 1007
QY 1665 attgacaacgccattctctgggccaactttggtatctctcacaacccacacacacacac 1724
QY 865 gtccccccctctgtttgttagtgcaatggggaatttagcgagaaatccatggagtc 924
DB ttt 1007
QY 1725 tgtgctggatcccgatctctcgtccagaggggcacatcagacaaagtctgcgcgcgcttc 1784
QY 925 cgtcatgctcatttgagagctctagagatggtctggccagaaacccattggaacccaag 984
DB ttt 1007
QY 1785 aaggagcgtgc-----ccagaagaacaaaggtcggaacccctctcgagcgagac 1832
QY 985 aggaacgatggtccctctcgcacaaagtcacagtagacagagatctctgggtaacattgac 1044
DB ttt 1007
QY 1833 accttccaggggtcccccaggtttccacgctccagttcagcgtatcatggagatcaac 1892
QY 1045 gttggc---aagrataccgcgagctcctcactggtggtgtagaaaggcgacaaagga 1101

DB 1893 caccgcaagaaggctggtgtactaccgtcgccacccggtggtgacccgcaacgaggggt 1952
QY 1102 ttgcgattgaacacgcagcatattgttcaatcccaaacccagggcagcaaaatttggttgag 1161
DB 1953 tactctatccagcactactgtcttcacagacgtctcacttccgacatgaagattgccaggag 2012
QY 1162 gagatctttggcccgctctgttcattaaagacgtttcaagacggaagaagagggccattgag 1221
DB 2013 gagatcttggctctgtctcactatccagaagtccaagatggtgaggtatcaag 2072
QY 1222 attgccaatgacacgacttatgg 1244
DB 2073 atcggcaantcgaccgactacgg 2095

RESULT 9
AAC78074
ID AAC78074 standard; cDNA; 3463 BP.
XX
AC AAC78074;
XX
DT 08-FEB-2001 (first entry)
DE Human cancer associated gene sequence SEQ ID NO:468.
XX
KW Human: cancer associated gene; cancer antigen; detection; cancer;
diagnosis; cytostatic; proliferative; vulnary; immunomodulator;
antidiabetic; antiallergic; antirheumatic; antithrombotic; antiviral;
antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;
dermatologic; neuroprotective; thrombolytic; coagulant; nootropic;
vasotrophic; antipapillary; antiangiogenic; gene therapy; inflammation;
immune disorder; haematopoietic cell disorder; autoimmune disorder;
allergic reaction; graft versus host disease; organ rejection;
haemostatic; thrombolytic; cardiovascular disorder; infection;
neurological disease; drug screening; ss.
XX Homo sapiens.
XX WO200055350-A1.
XX 21-SEP-2000.
XX
PD 08-MAR-2000; 2000MO-US05882.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUNA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587533/55.
XX P-PSDB; AAB43865.
XX
PT Novel isolated nucleic acids comprising sequences encoding peptides
useful for treating or diagnosing e.g. cancer -
XX
PS Claim 1; Page 1003-1004; 2352pp; English.
XX
CC AAC77607 to AAC78448 encode the human cancer associated proteins given
in AAB43398 to AAB44239. The proteins can have activities based on the
tissues and cells the genes are expressed in. Example of activities
include: cytostatic; proliferative; vulnary; immunomodulator;
CC antidiabetic; antiallergic; antirheumatic; antithrombotic;
CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatologic; neuroprotective; cardiac; thrombolytic; coagulant;
CC nootropic; vasotrophic; antipapillary; antirheumatic; The
CC polynucleotides and polypeptides can be used for preventing, treating or
ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
the present invention may be used to treat immune disorders from
or inhibiting the proliferation, differentiation or mobilisation of
immune cells, to treat disorders of haematopoietic cells, autoimmune

CC genetic constructions for the expression of a glycerol dehydratase
CC (GDH) and aldehyde dehydrogenase (ADH) capable of catalysing the
CC production of (3-HP) from (1); or (iii) carries a genetic construct
CC which expresses the dhaf gene from *Klebsiella pneumoniae* and a gene for
CC an ADH capable of catalysing the production of (3-HP) from (1). 3-HP is
CC a monomer, and is useful e.g. in the production of absorbable prosthetic
CC devices and surgical sutures or for incorporation into beta-lactams,
CC production of acrylic acid or formation of trifluoromethylated alcohols
CC or diols, polyhydroxyalkonates and copolymers with lactic acid.
CC Incorporation of genes encoding two enzymes makes the host organisms
CC able to produce (3-HP) from (1). The biotechnological method of
CC preparing (3-HP) is potentially cheaper than chemical synthesis. The
CC present sequence encodes the human aldehyde dehydrogenase ALDH2, which is
CC used in the exemplification of the present invention.
XX
SQ Sequence 1541 BP; 339 A; 418 C; 474 G; 310 T; 0 other;

Query Match 12.4%; Score 182; DB 22; Length 1541;
Best Local Similarity 48.3%; Pred. No. 3.3e-44;
Matches 701; Conservative 0; Mismatches 730; Indels 21; Gaps 6;

QY 12 gctcgcgaatacaagagtgaaactcttcacaaatgaattcgtctcctccaaaggggc 71
DB 66 gccagagctcttcgaaccagatttcaaaatgaatgaatgcagatgcgcgtcagca 125
QY 72 cagagattacgctcacgaacccgtggagcgaatccaccgttgcactgatgttcaagt 131
DB 126 gaaacattccaccctcgaacccgtccactggagaggtcattctgcaggtagctgaag 185
QY 132 ggcacacgcgcgcagtgctcagtgagtcagtagccgttcgtgcagcggtcaaaaagg 191
DB 186 ggaaca--ggagatgtggacaagcagtcgtgaaggccgcgcgttcctccagctgg 242
QY 192 cccatggaagaagttccaggtgcaacacgcgcgcgtgcatgcttaa---gttcgcgga 248
DB 243 ctacacttggcgcgcagtgacgacacacagcgcgcgtgctgaaacgcctggcga 302
QY 249 cctcgcgcgaagaacgcgcagagctcgtcgtcgtgagtcgctgccacgcgtagacc 308
DB 303 tctgatcagcgggacgcagactactcgttggcgttggagaccctggacaatggcaagcc 362
QY 309 gttgtcgtatcactcatttgcacattccaaacatggtctcgtgtttcgtactatgc 368
DB 363 ctatgctactctcactcgttggttgatgttgacatggtcctcaaatgctcgcgttattatgc 422
QY 369 aggcctggccgacaaagatccgcgaaagacatttccc---gagacaacgcgcaagccgaa 425
DB 423 cggctgggtgataaglacacgggaaacacatcccatggacgagacatttccagcta 482
QY 426 ttggcgttacgacgcagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 485
DB 483 cacacgcctgaacctgtgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 542
QY 486 tiacgtcgtcgaagatgacccgcgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 545
DB 543 gatcaagcatggaagctgggcccgccttggcaactggaaacgtggttgtgatgaagt 602
QY 546 ctcgagaaatcccgctggggt 605
DB 603 agctgagcagacacccctcacgcctctatgtggcaacctgatcaaggagctggcct 662
QY 606 cctcctcgtgagctgtagtctcactcgtgagcagagtgagcgggtgaa---gcattggc 662
DB 663 tccccctgggt 722
QY 663 gtcgcacatggacattgcgaagatcagcttccaaagatcgtcgcgcgttggcgcgcgcgc 722
DB 723 ctcccatgagatgtggacaagtggaattccacaggtccactcactgagattggccgta 782
QY 723 caagcaagcaacactcaagtccaacatgaagcgcgtcactctagactgggggaaagcc 782
DB 783 ccaggtgtgctgtggagcagcaactcaagagatgaccttggagctgggggggaaag 842

QY 783 aaccatgcttcaacgaagctcctctcgaacggcagtcggtggggaatcgcaaaagattt 842
DB 843 ccccaacatcatcatgtcagatgcgcgatatgattgggcccgtggaaacagggcccacttcgc 902
QY 843 ctcaaaattcgggcaaaatttgggtcccccctcctcgtttgtgtagtgcgaatggggaattt 902
DB 903 cctgtttcttcaacca-----gggccaagtcgtgtgcgcgcctccgcaccttcgtgca 956
QY 903 agcggagaaatccatggagtcgcgtcatgctcatttggaggtgagcgtgcagagatggtctgg 962
DB 957 ggagacatctatgatgatttgggtgcgcgagcgttgcgcgggccaagctcgggtgtg 1016
QY 963 ccagaaaccttggaaacccaagagcagcagtcgtcctcctcgcgaacagtcaccaaglacga 1022
DB 1017 cgggaaccccttgcagcaagc 1076
QY 1023 cagagcttgggttaacattgacgttggcgaag---gataccgcgcgcgcgcgcgcgcgcgcgc 1079
DB 1077 gaagatcctcgcgtcatcaacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1136
QY 1080 tggtagaaaggcgcgaagggattcgcgattgaaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1139
DB 1137 ggcatttgcgtgc 1196
QY 1140 aggcgcgaaaatttgggttggagagatcttttggccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1199
DB 1197 tggcatgacctcgcgaaggagagatcttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1256
QY 1200 gacggaagaagagccattgagattgcgaatgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1259
DB 1257 gacctagagagaggttgggtggagagccaaacatccacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1316
QY 1260 ttataccaaatctctcaacaggggtctcgcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1319
DB 1317 ctcccaaaaggatttggacaagggccaaattacatgctccagggccctccagcgcgcgcgcgcgcgc 1376
QY 1320 ctgcatacacttcccttttccgc 1379
DB 1377 ggggttcaactgctatgattgttgggcccagtcaccccttgggtgctacacagatgctc 1436
QY 1380 gggctcagggcagagcgttaggcgaagagggcctcaaggcgtacttggagccccaagaccat 1439
DB 1437 ggggagtgccgggagttgggc 1496
QY 1440 taatatccacgt 1451
DB 1497 cacagtcaaaagt 1508

RESULT 13
AAF18274

ID AAF18274 standard; DNA; 2071 BP.

XX AAF18274;

XX 14-MAR-2001 (first entry)

XX Lung cancer associated polynucleotide sequence SEQ ID 293.

XX Human; lung cancer associated protein; neuroprotective; cytosstatic;
XX cardioactive; immunomodulatory; muscular active; vulnerary;
XX gastrointestinal; nephrotropic; antiinfective; gynecological;
XX antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
XX proliferative disorder; wound healing; infectious disease; ds.

OS Homo sapiens.

XX WO200055180-A2.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000MO-US05918.

[illegible]

QY 1190 agacgttcaagacggaagaagagccattgagattgcccaatgacacgaccttatgggctag 1249
 Db 1416 tgaagttcagtgatgtggatgaggtgataaagaggcgacgagacgaagtacgggctag 1475
 QY 1250 cctcgtcatttatataccaatctctcaacaggggtctcctcgtgtctcgtcgcgctcgaga 1309
 Db 1476 cggcaggggttttcacgaagaatctggacacgcaaacaggggttcaagggtttgaaag 1535
 QY 1310 ccggtggtcgtcgtatcaacttccctttatccccgagacacaaactcgttttgcgcca 1369
 Db 1536 ctggtaccgtatgggttaactgtctcgaagttattgtagcaccattcggcggtt 1595
 QY 1370 tgaacaatcggggtcagcagagagctagggcgaagaagggtcgaaggcgtacttgagc 1429
 Db 1596 acaagatgagtggaatgggagagagaagacatatcacagttctcaataattacttgaga 1655
 QY 1430 ccaagac 1436
 Db 1656 tcaaggc 1662

Search completed: April 27, 2002, 00:40:59
 Job time: 18893 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2002, 23:02:13 ; Search time 7811.87 Seconds
(without alignments)
3921.782 Million cell updates/sec

Title: US-09-882-694A-4
Perfect score: 1464

Sequence: 1 atggtttcttcgcctgacga.....tccacgtcaacatagatga 1464

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sv.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htgo.inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB ID | Description |
|------------|-------|-------|--------|-------|-------------|
| ----- | | | | | |

| | 1 | 1464 | 100.0 | 1464 | 6 | AX076845 |
|----|-------|------|--------|------|-----------|----------|
| 2 | 270.4 | 18.5 | 1653 | 6 | UM0744668 | |
| 3 | 267.6 | 18.3 | 1488 | 8 | A43196 | |
| 4 | 267.6 | 18.3 | 1660 | 8 | CHCLAH3 | |
| 5 | 258.8 | 17.7 | 1659 | 8 | AF275347 | |
| 6 | 240.6 | 16.4 | 1751 | 8 | AB037421 | |
| 7 | 240.4 | 16.4 | 3021 | 9 | BC001619 | |
| c | 240.4 | 16.4 | 166207 | 9 | AL135785 | |
| 8 | 239.2 | 16.3 | 1485 | 6 | A43275 | |
| 9 | 239.2 | 16.3 | 1647 | 8 | AAALTA2 | |
| 10 | 238.2 | 16.3 | 6074 | 9 | HUMMTALD | |
| 11 | 238.2 | 16.3 | 79554 | 8 | NCB11B23 | |
| 12 | 232 | 15.8 | 2236 | 10 | BC020001 | |
| 13 | 232 | 15.2 | 2060 | 10 | MMU96401 | |
| 14 | 223.2 | 15.0 | 1563 | 10 | AF280404 | |
| 15 | 219.2 | 14.9 | 2048 | 3 | EBALDH | |
| 16 | 217.8 | 14.8 | 2024 | 10 | RATALDH | |
| 17 | 217.4 | 14.8 | 2372 | 10 | AF246711 | |
| 18 | 216 | 14.8 | 3263 | 10 | AF152359 | |
| 19 | 214.4 | 14.6 | 3390 | 10 | AF253409 | |
| 20 | 214.4 | 14.6 | 3390 | 10 | AF253409 | |
| 21 | 211.6 | 14.5 | 3442 | 9 | HSU07919 | |
| 22 | 209.8 | 14.3 | 1770 | 10 | AF001896 | |
| 23 | 209.8 | 14.3 | 2035 | 10 | AF001897 | |
| 24 | 209.8 | 14.3 | 2095 | 10 | AF001898 | |
| 25 | 209.6 | 14.3 | 1722 | 10 | MMRALDH2 | |
| 26 | 209.6 | 14.3 | 1855 | 8 | AB030939 | |
| 27 | 208.2 | 14.2 | 2012 | 10 | RNU79118 | |
| 28 | 208.2 | 14.1 | 9179 | 1 | PSP296087 | |
| 29 | 206.6 | 14.1 | 1806 | 10 | RATALDHA | |
| 30 | 203.2 | 13.9 | 2244 | 5 | AF339837 | |
| 31 | 202.2 | 13.8 | 2745 | 8 | AF260123 | |
| 32 | 202 | 13.8 | 11197 | 1 | AE009326 | |
| 33 | 202 | 13.8 | 11339 | 1 | AE008288 | |
| 34 | 201.6 | 13.8 | 2233 | 5 | AF315691 | |
| 35 | 201.6 | 13.8 | 2240 | 10 | RNU60063 | |
| 36 | 201.4 | 13.8 | 1630 | 9 | AB015226 | |
| 37 | 200.6 | 13.7 | 2745 | 8 | AF260124 | |
| 38 | 200.6 | 13.7 | 2745 | 8 | AF260125 | |
| 39 | 200.6 | 13.7 | 2745 | 8 | AF260126 | |
| 40 | 198 | 13.5 | 3217 | 8 | PAU40996 | |
| 41 | 196.8 | 13.4 | 1889 | 10 | RNADR | |
| 42 | 194.2 | 13.3 | 2804 | 8 | ENEALDA | |
| 43 | 193.2 | 13.2 | 1393 | 9 | AB015227 | |
| 44 | 192.8 | 13.2 | 3092 | 9 | AB015228 | |
| 45 | 191.4 | 13.1 | 1868 | 8 | AF269064 | |

ALIGNMENTS

| RESULT | 1 | AX076845 | Sequence 4 from Patent WO0105980. | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
|------------|--|--|-----------------------------------|---------|-----|--------|-----------------|
| AX076845 | LOCUS | AX076845 | Sequence 4 from Patent WO0105980. | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| DEFINITION | AX076845 | AX076845 | Sequence 4 from Patent WO0105980. | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| ACCESSION | AX076845 | AX076845 | Sequence 4 from Patent WO0105980. | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| VERSION | AX076845.1 | GI:13121519 | Sequence 4 from Patent WO0105980. | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| KEYWORDS | Exophiala spinifera. | Exophiala spinifera. | Sequence 4 from Patent WO0105980. | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| SOURCE | Exophiala spinifera. | Exophiala spinifera. | Sequence 4 from Patent WO0105980. | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| ORGANISM | Exophiala spinifera | Exophiala spinifera | Sequence 4 from Patent WO0105980. | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| REFERENCE | Duick,J.P., Maddox,J., Gilliam,J., Folkerts,O. and Crasta,O.R. | Duick,J.P., Maddox,J., Gilliam,J., Folkerts,O. and Crasta,O.R. | Sequence 4 from Patent WO0105980. | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| AUTHORS | Compositions and methods for fumonisins detoxification | Compositions and methods for fumonisins detoxification | Sequence 4 from Patent WO0105980. | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| TITLE | Patent: WO 0105980-A 4 25-JAN-2001; | Patent: WO 0105980-A 4 25-JAN-2001; | Sequence 4 from Patent WO0105980. | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| JOURNAL | Pioneer HI-Bred International Inc. (US) ; Curagen Corporation (US) | Pioneer HI-Bred International Inc. (US) ; Curagen Corporation (US) | Sequence 4 from Patent WO0105980. | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| FEATURES | Location/Qualifiers | Location/Qualifiers | Sequence 4 from Patent WO0105980. | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| SOURCE | 1.1464 | 1.1464 | Sequence 4 from Patent WO0105980. | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| BASE COUNT | 346 a | 402 c | 406 g | 310 t | | | |

ORIGIN

Query Match 100.0%; Score 1464; DB 6; Length 1464;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 atggtttcttcgctgacgaatacaagtgaaactcttcacaaatgaattcgtctcc 60
Db 1 ATGgtttcttcgctgacgaatacaagtgaaactcttcacaaatgaattcgtctcc 60
Qy 61 tcaaggggtccgagagattaaactcaacccggtgggacgaatccacccgttgccact 120
Db 61 TCAAGGGGTCCGAGAGATTAAACGCTCACCAGCCGCTGGGACGAATCCACCGTTGCCACT 120
Qy 121 gatgttcacgtggcacaacgcccgtatgtcgacagtcagtagcgttcggtcaggcg 180
Db 121 GATGTTCACTGGTGGCAACGCGCCGATGTGACAGTGCAGTAGCGGTTCGGTGCAGGCG 180
Qy 181 gtcaaaaggcccatggaagaagttcacaggtgcaacaacgcccgtgcatcttaag 240
Db 181 GTCAAAAGGCCCATGGAAGAAGTTACAGGTGCACAGGTGCACAGCGCGGTGATGCTTAAG 240
Qy 241 ttgcggacatcccgagaagaacgcgcgaagtcgctcgtcgtcgtcgtcgtcgtcgtc 300
Db 241 TTGCGGACCTCCCGGAGAAAGCGCGGAGAGTCTGCTCTGGAGTCGCTGCCACG 300
Qy 301 ggtagaccggtgtcgatgacatcatttcgacattccaaacatggtctcgtgtttcgc 360
Db 301 GGTAGACCGGTGTGATGATCATCTATTTGCACATTCCAACATGGTCTCCGTGTTCGC 360
Qy 361 tactatgcagctggcgcaagaatcgccggaagaccttcccgaggacaacggcaag 420
Db 361 TACTATGCAGCTGGCGGCAAGATCGCCGGAAGACCTTTCCCGAGGACAACGGCAAG 420
Qy 421 ccgaattggcgttacagccgatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
Db 421 CCgaattggcgttacagccgatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
Qy 481 tttctttacgtcgtggaagatagccccccgctcgtcgtcgtcgtcgtcgtcgtcgtc 540
Db 481 TTTCTTTACGTGCGTGGAGATAGCCCCGCTCGCCGCGGCTGCTCTTCATCTTC 540
Qy 541 aaagctcggagaaaatcccgctggcgcttctcgttctcgttctcgttctcgttctcgt 600
Db 541 AAAGCTCGGAGAAAATCCCGCTGGCGGTTCCTGGGCGCTCGCCGCGGCTGCTCTTC 600
Qy 601 ggaattccctcctgagtcgtgagttcctcactgagcagcagtcagcagtcagcagtc 660
Db 601 GGATTCCTTCCTGGAGTGTGCAAGTTCTCTACCTGGAGCAGTGCAGGTGAGCAATT 660
Qy 661 ggtgcacatggacattgcgaagatcagattcacaaatcgtcgcggtggccgcgc 720
Db 661 GGTGCGACATGCATTCGGAAGATCAGCTTCACAAGATCTGTCGGGTGGCCGGCC 720
Qy 721 gtcaagcaagcaactcaagtcacacatagaagcgtcactcactagaactgggggaaag 780
Db 721 GTCAAGCAAGCAACACTCAAGTCCAAATGAGCGGCTCCTCTAGAACTGGGGGAAAG 780
Qy 781 ccaaccatcgtctcaagaagctcctcgcagcagtcgaggggaaatcgcaaggat 840
Db 781 CCaaccatcgtctcaagaagctcctcgcagcagtcgaggggaaatcgcaaggat 840
Qy 841 ttctcaaaattcgggcaaaatttgggtccccccccctcgttctgtagtgaattgggaaat 900
Db 841 TTCTCAAAATTCGGGCAAAATTTGGGTCCCCCCCCCTCCTGTTGTAGTGAATGGGAAAT 900
Qy 901 ttacggagaaaattccatgagtcggtcagtcatttgagagctgtcagagatggctt 960
Db 901 TTACGGGAGAAAATTCATGGAGTCGCTCATGTGCTCATTTGGAGGCTGTCAAGATGGCTT 960
Qy 961 ggccagaacccattggaaccccaagagacgcatggtcccttcgtcagacagtcctcagtac 1020
Db 961 GGCAGAACCCATTGGAACCCCAAGAGACGCGCATGGTCCCTTCGTGACAAAGTCCCAGTAC 1020
```

```
Db 961 GGCCAGAACCCATTGGAACCCCAAGAGGACGCGCATGGTCCCTTCGTGACAAAGTCCCAGTAC 1020
Qy 1021 gacagagttttgggttaacattgacgttggcaagatataccgcgagctcctcactcgtt 1080
Db 1021 GACAGAGTCTTGGGTAAACATTACGTTGGCAAGGATACCGCGCAGCTCTCTCACTGGGTT 1080
Qy 1081 ggtagaaaaggcgacaaggattcgattgaaccacgacgatatattgtcaatcccaacca 1140
Db 1081 GGTAGAAAAGGGCGACAAAGGATTCCGGATTGAACCGACGATATTTGTCAATCCCAACCA 1140
Qy 1141 ggcagcaaaatttgggttgagagatctttggcccgctctgttccattgaagagcttcaag 1200
Db 1141 GGCAGCAAAATTTGGTTTGAGGAGATCTTTGGCCCGCTCTGTGTCATTAAAGAGCTTCAAG 1200
Qy 1201 acggaagaagaagccattgagattgccaatgacacgacttatgggtagcctcgttcatt 1260
Db 1201 ACGGAAGAAGAGGCGCATTTGAGATTGCCAATGACACGACTTATGGGTAGCTCGGTCAAT 1260
Qy 1261 tatacaaatctctcaacaggggtctcgtctcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1320
Db 1261 TATACCAATCTCTCAACAGGGGTCTCCGCTGTCTCGTGGCGCTCGAGACCGGTGGCGTC 1320
Qy 1321 togtacaaattccccctttatccccgagacacaaactccgtttggcgcatgaaacattcg 1380
Db 1321 TCGATCAACTTCCCCTTTATCCCGAGACACAACTCCGTTTGGCGCATGAAACAATCG 1380
Qy 1381 ggcctcaggcagagactaggcggaagggctcaaggcgtcacttgtagcccaagaccatt 1440
Db 1381 GCCTCAGGCGAGAGCTAGGCGGAAGGGCTCAAGCGGTACTTTGGAGCCCAAGACCAT 1440
Qy 1441 aatattccacgtcaacatagagtga 1464
Db 1441 AATATCCAGTCAACATAGAGTGA 1464

RESULT 2
UMU74468 LOCUS
DEFINITION Ustilago maydis cytosolic indole-3-acetaldehyde dehydrogenase
ACCSSION U74468
VERSION U74468.1 GI:1658174
KEYWORDS Ustilago maydis.
SOURCE Ustilago maydis
ORGANISM Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
REFERENCE 1 (bases 1 to 1653)
AUTHORS Basse,C.W., Lottspeich,F., Steglich,W. and Kahmann,R.
TITLE Two potential indole-3-acetaldehyde dehydrogenases in the
JOURNAL Eur. J. Biochem. 242 (3), 648-656 (1996)
MEDLINE 97175033
REFERENCE 2 (bases 1 to 1653)
AUTHORS Basse,C.W., Lottspeich,F., Steglich,W. and Kahmann,R.
TITLE Direct Submission
JOURNAL Submitted (14-OCT-1996) Institut fuer Genetik und Mikrobiologie,
Ludwig-Maximilian Universitaet, Maria-Ward Str. 1A, Muenchen,
Germany D-80638, Germany
FEATURES
source 1. .1653
/organism="Ustilago maydis"
/strain="FBI"
/db_xref="taxon:5270"
55_1548
/gene="iad1"
55_1548
/gene="iad1"
/EC number="1.2.1.3"
/function="converts indole-3-acetaldehyde to
indole-3-acetic acid in the presence of NAD(+)"
/note="cytosolic protein; belongs to the aldehyde
dehydrogenase family"
```

```

/codon_start=1
/product="indole-3-acetaldehyde dehydrogenase"
/protein_id="AAC49575.1"
/db_xref="GI:1658175"
/translation="MPTLNDLPLNGIKSTIQADLFNNKFPVALDCKTFTATNPSTGK
ETGQAEASAKVDLAVKAAAREAFETTWGNTPGDARGLLIKLAEIVANIDELAAI
ESLDNKFPIAKSFDAVAANRRYYGWADKHGKVMEDVTKRLYRTHPEIVGVC
EPLWQNPFLIMKAPALPATGNTIVLTAEQTPLSAIKMCELIYVAGPGVGVNVI
SFGPGVAGAAISOHMDIATFTGSTLVGRNIMKAAASTNLKVTLELGGKSNPIIFK
DADLQCAVWSAFGIFNFIHQCCAGCSRVVEESIYDAFMKKMTAHCKLQVGDPAKIA
NTFQGPVQSLOQYDRIMEYIESCKKDANALGVRKNGEYFIEPTFVDVPHDAKIA
KEIFGPVVPVFSKDEKDLIRANDSIYGLAAAFVSRDLSRAIEFTHAKLKGATVVMN
CYNQLIPGPPFGYKSGDGLREGLSALVNTNYTNKAVHNVLSQAPAI"
358 a 512 c 422 g 361 t

```

| Query Match | 18.5% | Score 270.4 | DB 8 | Length 1653 |
|-----------------------|--|-------------------|-----------|-------------|
| Best Local Similarity | 51.1% | Pred. No. 1.1e-55 | | |
| Matches 745 | Conservative 0 | Mismatches 676 | Indels 19 | Gaps 5 |
| QY 24 | caagagtaactctccatcaacaatgaatcgtctcccaagggtccgagagattaac | 83 | | |
| DB 102 | CCAGGCTGACTTGTTCATCAACACAGATTGTGCGCGCTCTTGACGGCAAGACTTTTGC | 161 | | |
| QY 84 | gtccacgaacctgggacgaatccaccgttgccactgatttcacgtggccaacgcgc | 143 | | |
| DB 162 | GACCATCAACCCCTCCACCCTTAAGGAGATTG---GTCAAGTCCCGAGGCTTCGGCAA | 218 | | |
| QY 144 | cgatgtcgacagtcagtagccgttcgggtcgcaaggcgtcaaaaaggcccatgaagaa | 203 | | |
| DB 219 | GGAGCTTGATCTCGTCAAGGTGCGCGAGAGGCAATCGAGACTACTTTGGGCGGAGAA | 278 | | |
| QY 204 | gttcacaggtgcacacgcgcggcgctcatgcttaagtctcgcgacactcgccgagaagaa | 263 | | |
| DB 279 | CACACCGGTCACCGCTCGTGGCAGACTCTGATCAAGCTGCCGAGCTGGTCGAGGCCAA | 338 | | |
| QY 264 | cgccgagaagctcgtcgttgagtcgctgctccaccgcggtagaccggtgtcgatgcac | 323 | | |
| DB 339 | CATTTCAGAGGTTGCTGCCATPCGAGTGCCTCGACAACGGCAAGGCGCTTCTCGATCGGTAA | 398 | | |
| QY 324 | tcatttcgaactccaaacatggtcctcggtgttcctactatcgactcaggctgggcccagaa | 383 | | |
| DB 399 | AAGCTTTGAGCTCGCTGCGCTGCTGCCAATCTCGATACATTGGTGGCTGGGCGGACAA | 458 | | |
| QY 384 | gatcgccggaagaccctttcccgaggacacgcgcaagccgaatt---ggcgttacagacc | 440 | | |
| DB 459 | GAACACGGCAAGGTCATGGAGGTCGACACCAACGCGCTCAACTACAGCGGCCACGAGCC | 518 | | |
| QY 441 | gatgggggtgtgctgagtagtcgcagctggaaacgcgaactttcttactcgtcgtctggaa | 500 | | |
| DB 519 | CATCGCGGTGTGTGGTCAGATCATCCCTTGGAACTTCCCTTGCTCATGTTTCGCTGGAA | 578 | | |
| QY 501 | gatagccccgcctcgccgcgcgctgcctctcatcttcaaaagcctcgagaaatccc | 560 | | |
| DB 579 | GCTCGGCGCTCGCTCGCCACAGTATACACGATCGTCTCAAAACCGCGGACGACACCC | 638 | | |
| QY 561 | gctgggggtctgggctcgtctctctcttcgcgagaagccggattccctcctggagtcgt | 620 | | |
| DB 639 | ACTGTACGCCATCAAGATGTGTGAGCTTATCGTCGAGGCGCGCTTCGCGCTGGTGTGT | 698 | | |
| QY 621 | gcagttcctcaatgg---agcacagtgtagcgggtgaagcattggcgctgcacatggacat | 677 | | |
| DB 699 | CAAGCTCATTTTCCGGTTTCGGTTCGCGTCCGCGTGGCGCCATCTCACAGCAATGGACAT | 758 | | |
| QY 678 | tcgaaagatcagcttcacaagatcgtcgcgcgtggtggcgccgttcgaagcaagcaaacct | 737 | | |
| DB 759 | TGACAGATCGCTTACCAGGCTCAACACTGGTGGGCGGCAACATATTGAAGGCTGTGC | 818 | | |
| QY 738 | caagtccaacataagacgcgcgtcaactctagaactgggggaaagccaacctctcttca | 797 | | |
| DB 819 | TTCCGACCAACTGAAGAGGTGAGCTCGAGCTCGGCGGCAAGTGCGCCCAACATATTCTT | 878 | | |

/db_xref="SWISS-PROT:P40108"
/translation="MTSVLETPHSGKYEQPTGLFINNEFKVGGKTFDVIINPSDES
VTVQHEATEKVDIVAAARQAFESRNLPTENRKLNNLNLEPKNFDLLAAVE
SLONGKATSKRVTSACASGCLRYVCGHADKTSKVIDTTPDTEYNYKKEPIGVCRSD
HSELEPLLMWANKIGIPAIACQNTVVLKTAQPTPLGLVAASLREAGSPGVINVISG
FKVGAALUSLHMDVDKVAFTGSTVVGRTILKAASNLKVKTLLELGSKNIVFEDA
DIDNALISWNFGIFFNHGQCCAGSRVVOESIDYKFKQKERAOKNVGDPPFAADT
FGQPOVSKYQFORIMEYIQAGKADAGATVETGSRKDGKGYFIETIFSNVTEDMKIVK
EETFGVCSIAFKTKEDAIKLGNASTVGLAAVHTKLNLTAEVSNALKAGTWWNT
YNTLHROMPFQGYKESGIGRELGEDALANYTQTKTVSIRLGDALFG"
BASE COUNT 371 a 502 c 440 g 347 t
ORIGIN

Query Match 18.3%; Score 267.6; DB 8; Length 1660;
Best Local Similarity 53.0%; Pred. No. 5.4e-55;
Matches 766; Conservative 0; Mismatches 654; Indels 26; Gaps 8;

QY 17 acgaatacaagagtgaaactcttcaacaataatgaattctctccctcaagggtccgaga 76
DB 110 ACGAGACGCGCGCGCTCTTCAACAACAGAGTTCGTCAAGGGCCCAAGAGGCAAGA 169
QY 77 gattaaagctcaagaacccgtggagcaaatccacccgttgcactgatgttcacgtggcca 136
DB 170 CCTTCGATGTCATCAACCCCTCCGACGA--GAGCGTGATCACCCAGGTCCAGAGGCCA 226
QY 137 acgagccgagtgatgcagcagtagcagcgttcggttcgagcggtgcaaaagggcccat 196
DB 227 CCGAGAAAGGATGTCAGATCGCGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 283
QY 197 ggaagaagttcaagtgagcaaacgagcgcggtgcagtccttaagttcgcggaccccgccg 256
DB 284 GGAGCTGGAGACACCCGAGAACCGTGGCAAGCTGCTCAACAACCTCGCCAAACCTGTTCG 343
QY 257 agaagaacgcgagagagtcgctgtctgtgagtcgctgcaccccgctagacgctgttoga 316
DB 344 AGAAGAACATGACCTCTCTGCGCTGAGTGGCTGACACACGCGCAAGGCGCACTTCCA 403
QY 317 tgatcactcttgcagatccaaacatggtctcgtgttcgtactatcagcagctggg 376
DB 404 TG---GCAAGGTGACATGACATGTCGCTGCGGTGCGCTCAGATACGTTACGTTGGG 460
QY 377 ccgacaagatgcgcggaagacclttcccgagagacaacggcgaagcgaattggcgcttac- 435
DB 461 CGGACAAGATCAACGGCAAGGTGTCAGACTACGCGCGCGACACTTTCAACTACGTCAGA 520
QY 436 --gagccgatgggggt 493
DB 521 AGGAGCCCATGTTGGCGGTGATGATCATTCCTGGAACCTTCCCTTCTCATGTGGG 580
QY 494 gctggaagatagcccgccctgc 553
DB 581 CATGGAAGATCGCGCGCGCGCATTCGTTGCGGTAAACACTGCTGCTGCTGGAAGACTGCTGAGC 640
QY 554 aatcccgctggcgcttggcgctgcctctctctctctctctctctctctctctctctctct 613
DB 641 AGACCCCTCTGT 700
QY 614 gagtgcgtgcagttcctcactggagacaga---gtgacgggtggaagcattggcgtgcgaca 670
DB 701 GTGTGCATACGTCATTCGGGTTTCGGCAAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 760
QY 671 tggacattgcgaagatcagcttcaaatctctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 730
DB 761 TGGACGTCGACAAAGGTGGGCTTCACCGGTTCCACCGGTTCCACCGGTTCCACCGGTTCC 820
QY 731 caacactcaagtcacatgaagcgcgtcactctgaagcgtggggaagcgaacacatcg 790
DB 821 CTGTGCTCTTCCAACTTGAAAGGTACCCCTCGAGTCCGTTGGCAAGTACCAACAACA 880
QY 791 tcttcaacgaagctcctctcgaacggcagtcggtggggaatcggcgaagatttctcaaat 850
DB 881 TTGTCTTCGAGGACCGCGATATTGACACGCGCATCTCATGTTGGTCAACTTCGCTATCTTCT 940

QY 851 tcggggcaaatgttggtccccccctcctgttctagtgcaatggggaatttagcggaga 910
DB 941 TCAACACAGGGCCAGT 993
QY 911 aattccatgagtcgcgtcattgttggt--gctgtcagagatggttggccagaga 969
DB 994 ATCTACGACAAGTTCGTCCAGAAGTTCNAGGAGCGCGCACAGAGAACGTTGTTGGCGAC 1053
QY 970 ccattggaacccaagagacgcagtcctcctcgtgcagaagtcctcagtcacagagtc 1029
DB 1054 CCCTTCGCGCGCGCACATTCACAGGTCTTCAGGTTTCNAGGTTTCAGTTCGACCCCAT 1113
QY 1030 ttgggttaacattgagttggcaaggatacc---gcgcagctcctcctcctggttggtaga 1086
DB 1114 ATGAGTATACATCCAGGCGCGCAAGGACGCGGTGCCACCTCGAGACCGGTGGAAGCCGT 1173
QY 1087 aaggcgcaagaggttcgcgattgaaccgacgatatattgtcaatcccaaacaccagcgac 1146
DB 1174 AAGGTGACAAGGCTACTTCATTTGAGCCCAACCATCTTCTCCAACTCAGAGGACATG 1233
QY 1147 aaaattgtgttgaggagatttggcccgcttcttccatttaagcgttcaagacggaa 1206
DB 1234 AAGATCGTGAAGAGGAGATCTTCGGCCCGCTCTGCTCGATGCCAAGTTCAGACCAAG 1293
QY 1207 gaagagccattgagattgccaatgacacgacttatgggctagcctcgttcattatacc 1266
DB 1294 GAGGACGCCATCAAGCTCGGCAACGCCAGCACATACAGGTCTCGCCGCCGCGTCCACAC 1353
QY 1267 aatctctcaacaggggtctccgtctcgtcgcgcgtcgcagaccggtggcgtctcgatc 1326
DB 1354 AAGAACCTCAACACCCGCTCGAGGTCTCCAACTGCTCAAGGCGGCGACCGTCTGGGTCT 1413
QY 1327 aacttcccttattcccccagagacacaaactccgtttggggcgatgaaacaactgggctca 1386
DB 1414 AACACTTACAACACCCCTCCACCCAGATCCGCTTGGTGGCTTACAGAGTCCGCTATT 1473
QY 1387 ggcagagctagcggaagagggctcaagcgtacttggagcccaagaccattataatc 1446
DB 1474 GGCGCTGAGCTCGGCGAGGATGGTGGCCCACTACTACTCAGACCAAGACTGTTAGCAT 1533
QY 1447 caegtc 1452
DB 1534 CGTCTC 1539

RESULT 5
AF275347
LOCUS Cladosporium fulvum aldehyde dehydrogenase (aldh) mRNA linear JUN 13-JUL-2000
DEFINITION cds.
ACCESSION AF275347
VERSION AF275347.1 GI:9082282
KEYWORDS Cladosporium fulvum.
SOURCE Cladosporium fulvum
ORGANISM Cladosporium fulvum
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes et
Chaetothyriomycetes incertae sedis; Mycosphaerellaceae; mitosporic
Mycosphaerellaceae; Cladosporium.
REFERENCE 1 (bases 1 to 1659)
AUTHORS Coleman,M., Henricot,B., Arnau,J. and Oliver,R.P.
Starvation-induced genes of the tomato pathogen Cladosporium fulvum
are also induced during growth in planta
Mol. Plant Microbe Interact. 10 (9), 1106-1109 (1997)
JOURNAL 98052127
MEDLINE 9390425
REFERENCE 2 (bases 1 to 1659)
AUTHORS Segers,G.C., Bradshaw,N., Archer,D., Blissett,K. and Oliver,R.P.
Alcohol oxidase is a novel pathogenicity factor for Cladosporium
fulvum but aldehyde dehydrogenase is dispensable
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1659)
AUTHORS Segers,G.C.

| TITLE | Direct Submission |
|------------|--|
| JOURNAL | Submitted (05-JUN-2000) CAB, University of Maryland, 5115 Plant Sciences Bldg., College Park, MD 20742, USA |
| FEATURES | Location/Qualifiers |
| source | 1..1659 |
| | /organism="Cladosporium fulvum" |
| | /strain="Cf4" |
| | /specific_host="Lycopersicon esculentum Mill." |
| | /db_xref="taxon:5499" |
| gene | 1..1659 |
| | /gene="aldh" |
| CDS | 61..1554 |
| | /gene="aldh" |
| | /note="psi-9; ALDH" |
| | /codon_start=1 |
| | /product="aldehyde dehydrogenase" |
| | /protein_id="AAF82789.1" |
| | /db_xref="GI:9082283" |
| | /translation="MSLSVLETPSAGKYEOPTGLFINNEWKPVSGKFPDVINPTDSESVICQVAEADKQVDIAVAAAKAFCEPHQETPENRGLLVKLAELEFNKLDLFASV KALDNGKAFNMKVDIGPATAGTGVYVKTAEQTPLGAYVAANLIKKEGPPGVGVGQITIPWNPFLLMWAKIKGIVATGTVVYIKTAEQTPLGAYVAANLIKKEGPPGVGVGQIFGFKAGAAITASHMDIDKVAFTGSTVTGRTILKAAAGSNLKKVTLLEGKSPNIVFDDADIDNAISWVNFQIFGFIHNGRCOCAGSRIYVQEGIDYKFLQRFKRAQKNVGVGPPFAADTFQGPQVQVQFQIMGYIQACRGDAGTVEIGNRKGDGYFIEPTIPANVTEDMKIMQEEIFGPVCSISFKETKEDIAKVGNNATPPGLAAAVHTRNLTAIEVSNALRAGTVVVM TYNTLHQHPFGFGKSGIGRELGEDALANYTQTKTVSIRLGDALFG" |
| BASE COUNT | 381 a 496 c 440 g 342 t |

| Query Match | 17.7% | Score 258.8 | DB 8 | Length 1659 |
|-----------------------|----------------|--|-----------|-------------|
| Best Local Similarity | 52.8% | Pred. No. 7.8e-53 | | |
| Matches 763 | Conservative 0 | Mismatches 647 | Indels 36 | Gaps 8 |
| QY | 17 | acgaatacaagagtgaaactcttcatcaacaatgaattcctctccaaaggggtccgaga | 76 | |
| Db | 104 | ACGAGCAGCCAACTGGCCTTTTCATCAACAATGAGTGGGTGAAGCCAGTCAGCGCAAGT | 163 | |
| QY | 77 | gattaacgctcaagaacccgctggagaaatccacggtggccactgatgttcacgtggcca | 136 | |
| Db | 164 | TCTTCGATGTATCAACCAACCCAGCAGAGT---CCGTATCTGCACAAGTCGACAGGCGG | 220 | |
| QY | 137 | acgvgcgcgatgtcgacagtgacgtagccgtctcgggtgcagcgcggtcaaaaagggccat | 196 | |
| Db | 221 | ATGAGAAGGATGTCGACATCCCGCTCGCGCGCCGCCAAGGCCCTTGA---GGGCCCAT | 277 | |
| QY | 197 | ggagaaggttcacagggtgcacaacgcgcggtgcattgcttaagttccgcgacctcgccg | 256 | |
| Db | 278 | GGAGACAAGAGACACCCAGAGAACCGCGGCAAGCTTCCTCGTGAAGTCGCCGAGCTCTTCG | 337 | |
| QY | 257 | agaagaacgcgcgagaagctcgctcgctggaagtcgctgccacgcggtagaccgggtcga | 316 | |
| Db | 338 | AGAAAGACCTCGACCTCTTCGCCTCCGCTCAAGGCGCTCGACAAGC---GAAAGCGGTTC | 394 | |
| QY | 317 | tgatcaactatttcgaattcccaaaacatggctcgtgttctgctactatgcagggctggg | 376 | |
| Db | 395 | ACATGGGCAAGGTTGACATTTGGATGTGCGCTGGCTGCCCTCCGATACTACGGTGGCTGGG | 454 | |
| QY | 377 | ccgacaagatgcgcggaagacacttcccaggagacaacgcgaacgaattggcgt---t | 433 | |
| Db | 455 | CCGACAAGATTGAGGGCAAGGTCGCGACACCACTCCAGATACCTTCAACTACATCAGA | 514 | |
| QY | 434 | acgagcggaatgggggtgtgtgctggtattgccagctggaaacgcgactttctttaagtcg | 493 | |
| Db | 515 | AGGAGCAATTCGGTGTGTGCGGTGTCAGATCATCCCATGGAACTTCCCACTCCTCATGTGGG | 574 | |
| QY | 494 | gctggaagatagccccgcgcctcgccgcgggtgctctctcatcttcaagcctcggaga | 553 | |
| Db | 575 | CATGGAAGATTGCTCCAGCCGTTGCCACGTGTAACACCGTTGTTCATCAAGACATGCTGAGC | 634 | |
| QY | 554 | aatccccgcgtggcggttcttggccctgcctctcttcgagaagcggattccctctcg | 613 | |

ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.

REFERENCE
AUTHORS Li, X., Nakazono, M., Tsutsumi, N. and Hirai, A.
TITLE Molecular and cellular characterizations of a cDNA clone encoding a novel isozyme of aldehyde dehydrogenase from rice
JOURNAL Gene 249 (1-2), 67-74 (2000)
MEDLINE 20293048
REFERENCE 2 (bases 1 to 1751)
AUTHORS Nakazono, M.
TITLE Direct Submission
JOURNAL Submitted (23-JAN-2000) Mikio Nakazono, The University of Tokyo, Laboratory of Plant Molecular Genetics, Grad.Sch.Agr.Life Sci.; Yayoi 1-1-1, Bunkyo-ku, Tokyo 113-8657, Japan
(E-mail: anakazon@mail.ecc.u.tokyo.ac.jp, Tel: +81-3-5841-5074, Fax: +81-3-5841-5183)

FEATURES
source Location/Qualifiers
1..1751
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="EST clone R1467"
63..1571
/gene="ALDH1a"
63..1571
/gene="ALDH1a"
/codon_start=1
/product="cytosolic aldehyde dehydrogenase"
/protein_id="BAA96794.1"
/db_xref="GI:8574437"
/translation="MAAANGDSKGFVPKLEIKFLINRFDVAVSGKTFETRDP
RTEVIAKIDGKADIDLAVKAAEAFDHGPRMSGFARGRLHKFADLVEQHVEE
LAALPTVAGKLIVAGADLLRYPGAGNLLRYAGAAQKVHGETLKMARPCHYTLKEPV
GVVGHVFWNTPTTFKPSALAAAGCTMVVPAEOTPLSALFYAHLAKLAGVDPGV
LNVVPGFPTAGAAISSHMDIKVSTGTEVGRVLMEEAASNLKPYLSLELGGKSPV
IVFDDADLTAVNLVHMASYTNKGLICVAGSRVLYVQEGYDAFYKATEMAKKSVVGD
FENPRVHGGPQIDKQYKILKYIDIGKREGATLVGKPGENGYYLEPTFYDVRKE
EMSAIEELFGFPMALMFKTVEAIIQRANSTRYGLAAGIVTKNIDIVANTVSRIRAG
AIIWICYLGFDPDVPFGKMSGFGKDMDALEKLYLTKAVVPLYNPL"
BASE COUNT 406 a 518 c 512 g 315 t

Query Match 16.4%; Score 240.6; DB 8; Length 1751;
Best Local Similarity 50.9%; Pred. No. 2.3e-48;
Matches 731; Conservative 0; Mismatches 684; Indels 20; Gaps 6;

QY 32 aactcttcacacaatgaattcgtctctccaaaggggtccgagagattaaacgctcaaga 91
DB 130 AGCTCTTCATCAATGCGCGCTTCGTGCGACGGCTCTCCGCGCAAGACATTCGAAACCCGNG 189

QY 92 accctggagacaatccaccgttgccactgatttcaactggtgccaacgcgcatgtcg 151
DB 190 ACCCGCGCACCGCGCGAGGTTCATCGGCA---AGATCGCGGAAGGAGACAAAGCGCATCG 246

QY 152 acagtgcagtagcgtcttcgtgacgcggtcaaaaagggcccatggaagaagttaacag 211
DB 247 ACCTCCCGGTGAAGCGCGCGGAGGCGCTTCGACCATGCGCCCTGGCCCAAGATGTCG 306

QY 212 gtgcaacaacgcggtcgatgcttaagtctgcggacctcgcgagaagaacgcgaga 271
DB 307 GCTTTTCGAGGGGAGAGATCCTGCACAAAGTTCGCGGACCTGGTGGAGCAGCAGCTGGAGG 366

QY 272 aactcgtcgtctgaggtcgctgccaccggttagccggtgctgcatgcatcatttgg 331
DB 367 AGCTGCGCGCTGGACACGGTGGACCGCGGCAAGCTGTTCGGATGGGGAAGCTCGCTG 426

QY 332 acattccaacatggtctcgtgtttcgtactatgcatgagctgggcccgaagaatcgccg 391
DB 427 ACATCCCGCGCGCGGAACCTGCTCCGGTACTACGCGCGCGCGGACAAAGGTGCACG 486

QY 392 gaaagacatttcccgagagacaa---cggcaagccgaattggcggttacgagcgcatggggg 448
DB 487 GCGAGACGCTCAAGATGGGCGCGCCATGCCAGGGGTACACGGGTCAAGAGGACCGCTCGCG 546

QY 449 tgggtgctggtattgccagctggaacgcgactttttttacgtcggtcggaagatagccc 508
DB 547 TCGTGGCGCACATCGTGGCGGTGGAACATCCCCACACCACTGTTCTTCTTCAAGGCCAGCC 606

QY 509 cgcctctcgcgcgggtgctctcttcatcttcaaaagcctcgagaaatcccgcgtggggcg 568
DB 607 CGGCGCTCGCGCGGCTCACCATGCTGCTCAAGCGCCGCGAGCAGACCCCTCTCCG 666

QY 569 ttctggcctcgtctctctcttcgcaagccggtattccctcctggtgagtcgagtcctc 628
DB 667 CGCTCTTCTACGCCCACTCGCAACGTTGCCGGCTCCCGACGCGGTGCTCAAGCTCG 726

QY 629 tcactggagacagagtga---cgggtgaagcattggcgctgcacatggacattgcaaga 585
DB 727 TCCCCGCTTCGGCGCCACCGCGCGGCTATCTCTCTCCACATGGACATTTGACAAG 786

QY 686 tcagcttcaagaatctctcgcggtggtggtggtggtggtggtggtggtggtggtggtggt 745
DB 787 TGAGCTTACCGGCTCGACGGAGGTGCGCGGCTGCTGATGAGGGCGGCGGCAAGAGCA 846

QY 746 acatgaagcgtcactctagaaactgggggaaagcccaaccatcgcttctcaacgaagctc 805
DB 847 ACCTGAAGCCGCTCGCTCGAGCTGGGTGGCAAGTCTCCGGTCACTCGTGTTCGACGACG 906

QY 806 ctctgaaacgcgctcggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 865
DB 907 CCGACCTCGACACGGCGCTGAACCTCGTCCACATGCGCTCTTACACCAACAGGGTGAGA 966

QY 866 tccccccctcttctgtgtagtgcattggggaatttagcggagaatttccatggagctc 925
DB 967 TCTGCTGCCGCTCGCAAT-----CTATGTCAGGAGGGGATCTACGATGCGTTC 1019

QY 926 gtcatggtcatttgaggtctcagagatggtctggtggtggtggtggtggtggtggtggtggt 985
DB 1020 GTGAAGAAGCGACCGAGATGCGCAAGAAATCGTGGTGGAGATCCGTTCAACCCGCGCA 1079

QY 986 ggtac-gcatggttccctctgcaacagtcctcagtagtcagacagagcttgggttaactgac 1044
DB 1080 GTTATCAAGGCCCTCAGATTGACAGGAGCAATACGAGAAATCCTCAAGTACATCGAC 1139

QY 1045 gttggcaag---gataccgcgagctcctcactggtggtggtggtggtggtggtggtggtggt 1101
DB 1140 ATCGGTAAAGCGGAGGCGGACGCTTGGTTCACCGGAGGAAAGCCCTGCGCGGAAACGSG 1199

QY 1102 ttgcgattgaaccgcagatatttgcatacccaaccagcgagcaaaaatttggtttggag 1161
DB 1200 TACTACATCGACCCACCATCTTCAGGAGCTCAAGGAAGAAATGTCGATCGGCAAGAG 1259

QY 1162 gagatatttggcccgctgttgcattaaagacgttcaagacggaagaagagccattgag 1221
DB 1260 GAAATCTTCGGCGGCTGATGCGCCCTCATCAAAATTCAGACGCGTTGAGGAGCGCATCCAG 1319

QY 1222 attgcaatgacacgacttattggtcgtcgtgctcatttaccacatctctcaacag 1281
DB 1320 AAGGCGAACAGCACCGGTCACGGCTGGGTGCGCGATAGTACCCAAAGAACATCGACGTC 1379

QY 1282 ggtctcgtgctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1341
DB 1380 GCGAACACGTTTCGGCTCGATCCGGCGGCGAGGGGCAATCTGGATCAATTCGTACCTCGCG 1439

QY 1342 cccgagacacaaactcgttttgccgcaagaaatcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1401
DB 1440 TTCGACCCCGAGCTGCGCTTCGGGGGTACAAGATGACGGGCTTCGGCAAGACATGGGC 1499

QY 1402 gaagaaggctcaaggcgttacttggagcccaagacattatataccagctcaaca 1456
DB 1500 ATGGAGCCCTTGAGAAAGTACCTCCACACCAAGCGGTGTCACCCCTCTCTTACA 1554

QY 1076 qcgttgataaaggcgcaaggattccgattgaaacggagcagatattgtcaatccca 1135
 Db 1250 GCGAGAGCCTTCGGGGAGCGTGGTTCCTTCAATCAAGCTACTGTCTGTGGTGGCGTGC 1309
 QY 1136 aaccaggcaaaaattgtttgagagatctttggccccctcttcttcattagaagct 1195
 Db 1310 AGGATGACATGAGAATTGCCAAGAGGAGATCTTTGGGCTGTGCAGGCCCTGTTCAGT 1369
 QY 1196 tcaagcgaagaagagcattgacattgccaatgacagactatggctagctcgg 1255
 Db 1370 TCAGAAGATTGAGGAGGTGTGAGAGGGCCCAACACACAGGATGCGCTGGCTGCGG 1429
 QY 1256 tcattataccaaatctcaacagggtctcgtctcgtcgcgcgtcgagaccgtg 1315
 Db 1430 CTGTGTTCACTCCGGATCTGGACAAGCCATGTACTTCAACCCAGGCACCTCCAGCCGGA 1489
 QY 1316 gcgtctgatcaactccctttatcccgagacacaaactcgtttggcgcatgaac 1375
 Db 1490 CCGTGTGGGTAACACTCAACATCGTCACCTGCCACACGCCATTTGGAGGGTTTAAGG 1549
 QY 1376 aatcggtctcagcagagctagtcgaagaagggtcgaagcgtacttgagcccaaga 1435
 Db 1550 AATCTGAAACGGGAGGAGCTGGTGAGATGGGCTTAAGGCTACACAGAGGTAAGA 1609
 QY 1436 ccattaataccacgt 1451
 Db 1610 CGGTCAACATCAAGGT 1625

RESULT 8
 ALL35785/c
 LOCUS
 DEFINITION Human DNA sequence from clone Rp11-113024 on chromosome 9p11-13.3
 Contains a novel gene for a protein similar to insulin-like growth factor binding protein, the ALDH1B1 gene for aldehyde dehydrogenase 1 family, member B1 and CpG islands, complete sequence.
 ALL35785
 AL135785.10 GI:10803183
 HTG; aldehyde dehydrogenase; ALDH1B1; ALDH5; ALDHX; CpG island.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 166207)
 AUTHORS Laird,G.
 TITLE Direct Submission
 JOURNAL Submitted (04-DEC-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 On Oct 16, 2000 this sequence version replaced gi:10086008.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
 En: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/Celegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr9
 This sequence is the entire insert of clone Rp11-113024 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. Rp11-113024 is from

the library RPCI-11.1 constructed by the group of Pieter de Jong.
 For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pBACe3.6.
 Location/Qualifiers

FEATURES

source

1..166207
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosomes="9"
 /map="p11-13.3"
 /clone="Rp11-113024"
 /clone_lib="RPCI-11.1"
 8..568
 /note="match: GSS: Em:AQ341800"
 12..429
 /note="match: GSS: Em:AQ047065"
 85..242
 /note="MUT1J repeat: matches 103..261 of consensus"
 502..624
 /note="MLT1B repeat: matches 1..121 of consensus"
 948..1148
 /note="MIR repeat: matches 49..262 of consensus"
 1124..1174
 /note="L2 repeat: matches 2699..2743 of consensus"
 1734..1791
 /note="2 copies 29 mer 96% conserved"
 2438..3163
 /note="MER4A repeat: matches 1..660 of consensus"
 3460..3529
 /note="L2 repeat: matches 2641..2710 of consensus"
 3802..3909
 /note="MIR repeat: matches 19..123 of consensus"
 4126..4528
 /note="match: GSS: Em:AQ175777"
 4135..4559
 /note="match: GSS: Em:AQ407389"
 4243..4374
 /note="FRAM repeat: matches 1..132 of consensus"
 4556..4802
 /note="L2 repeat: matches 2245..2496 of consensus"
 5095..5651
 /note="MER6A repeat: matches 1..605 of consensus"
 complement(5300..5808)
 /note="match: GSS: Em:AQ175541"
 5829..6461
 /note="match: GSS: Em:A2517757"
 5917..6000
 /note="HERVL repeat: matches 2094..2185 of consensus"
 6067..7341
 /note="HERVL repeat: matches 564..1941 of consensus"
 7381..7903
 /note="match: GSS: Em:AQ387157"
 7634..7923
 /note="AluJo repeat: matches 1..301 of consensus"
 7955..8266
 /note="AluX repeat: matches 1..310 of consensus"
 8272..8573
 /note="AluJb repeat: matches 1..299 of consensus"
 8784..8928
 /note="L1ME repeat: matches 812..955 of consensus"
 8939..9235
 /note="L1ME repeat: matches 310..621 of consensus"
 9285..9698
 /note="match: GSS: Em:AQ428063"
 10632..10813
 /note="AluJo repeat: matches 98..272 of consensus"
 11025..11097
 /note="L1MC/D repeat: matches 5735..5809 of consensus"
 11101..11136
 /note="18 copies 2 mer tt 83% conserved"
 12555..12654
 /note="MIR repeat: matches 22..131 of consensus"
 12571..13348
 misc_feature

```
/note="CpG island"
/evidence=not_experimental
Join(12875..13334,23096..23205,23946..24062,25750..25899)
/Join("Bali3024.1"
/Join(12875..13334,23096..23205,23946..24062,25750..25899)
/Join("Bali3024.1"
/note="match: cDNAs: Em:AB006141
match: proteins: Tr:O88812 Tr:Q16270"
/codon_start=1
/evidence=not_experimental
/product="Bali3024.1 (similar to Insulin-like growth
factor binding protein)"
/db_xref="GI:17384405"
/translation="MPLSLILPLILLLPLPLPLSLGIRVGGRRPKCGPCRPE
GCPAPCPAPGISALDCGACRLGAGSCGGRAGRGPGILVCAQAAAGAPEG
TGLCVCAQGVCGSDGRSPVSCALRLRARHTPRAHPGHLHAKRDPCEAFAPVVPV
PRSHVNTVGAQVGLSCFVRVPTVITWKEVTKSPEGTQALEELPGDRHVNIAQVRGG
PSDHEATWILINLURKEDEGVYQCHAAANMGEASHSTVTVLDLSKTRSFHPAPDD
RM"
12890..12940
/note="17 copies 3 mer ctg 86% conserved"
13062..13139
/note="26 copies 3 mer gcg 66% conserved"
13692..13787
/note="4 copies 24 mer 85% conserved"
14013..14091
/note="MER69 repeat: matches 1..78 of consensus"
15994..16286
/note="AluJo repeat: matches 3..303 of consensus"
16338..16403
/note="MIR repeat: matches 76..146 of consensus"
16444..16747
/note="AluJo repeat: matches 1..300 of consensus"
17142..17406
/note="AluJo repeat: matches 34..298 of consensus"
17407..17457
/note="17 copies 3 mer gga 92% conserved"
18914..18997
/note="MIR repeat: matches 112..183 of consensus"
19236..19756
/note="MER68A repeat: matches 1..561 of consensus"
19842..19967
/note="MIR repeat: matches 34..188 of consensus"
20015..20199
/note="MIR repeat: matches 7..241 of consensus"
20328..20622
/note="AluJo repeat: matches 10..299 of consensus"
20641..20780
/note="L1MEL repeat: matches 6017..6150 of consensus"
22762..22896
/note="MIR repeat: matches 5..140 of consensus"
23235..23264
/note="15 copies 2 mer gt 100% conserved"
24244..24733
/note="L1TID repeat: matches 8..505 of consensus"
26799..27098
/note="AluJ repeat: matches 1..299 of consensus"
Complement(27151..27534)
/note="match: GSS: Em:AQ131150"
Complement(27159..27539)
/note="match: GSS: Em:AQ128717"
27545..28093
/note="match: GSS: Em:AQ763725"
27546..28104
/note="match: GSS: Em:AQ768584"
28476..28558
/note="MIR repeat: matches 65..143 of consensus"
28848..29167
/note="AluJo repeat: matches 1..302 of consensus"
29434..29574
/note="MIR repeat: matches 87..242 of consensus"
29929..30009
```

```
misc_feature
/note="MIR repeat: matches 77..155 of consensus"
Complement(30372..30771)
/note="match: STR: Em:G23339"
Complement(31296..31713)
/note="match: GSS: Em:AQ683842"
31311..31411
/note="L2 repeat: matches 2319..2418 of consensus"
Complement(31523..31902)
/note="match: GSS: Em:AQ130619"
33086..33279
/note="L1TID repeat: matches 63..265 of consensus"
33439..33742
Query Match 16.4% Score 240.4; DB 9; Length 166207;
Best Local Similarity 49.8%; Pred. No. 3.2e-48;
Matches 725; Conservative 0; Mismatches 716; Indels 15; Gaps 4;
QY 5 ttctttcgctgacgaatacaagagtgaaactcttcttcatacaaatgaattcgtctctcca 64
DB 41465 TTCTGAACCCGACATCCCTACACCGCTGTTCATCAACAATGAATGGCAAGATGCAG 41406
QY 65 aggggtccgagagattaaacgtcacgaacccgtgggacgaataccaccgttgccactgatg 124
DB 41405 TCAGCAAGAAGACCTTCCCGACGGTCAACCTTACCACCGGGGAGGTCTATTGGCAGCTGG 41346
QY 125 ttcacgtggccaaacggcgccgatgtcgacagtgacagtagcagcttcgttcgagggcggtca 184
DB 41345 CTGAAGGTGACCGGGCTGATGTGGATCGGGCGGTGAAGACGACGGCGGGAAGCTTCCGGC 41286
QY 185 aaaaagggccatggaagaagtccaaggtgcacaacgcgcgcgtgcagtcttaagtctcg 244
DB 41285 TGGGTGCCCATGCGCGGATGGATGCTCTGAGCGGGCGCGGTGTGTAACCGCTGG 41226
QY 245 cggacctgcgcagagaacgcgcgagaagctcgtcgtctggtgagtcgtgcacacccggtca 304
DB 41225 CAGACCTAGTGAGCGGGATCGAGTCTACTTGGCTCTACATCGAGACCTTGGACAATGGGA 41166
QY 305 gaccggtgcgatgactcatttcgacattcccaaacatggctccgtgttcgtact 364
DB 41165 ACCCTTCAACAGTCTTACGCTTGGACTTGGATGAGTGCATCAAGGTGTATCGGTACT 41106
QY 365 atcgaggtggcgacgaagatcgcggaagaacctttcccgaggaacacgcgcagccga 424
DB 41105 TTGCTGGCTGGCTGACAAAGTGGCATGGCAAGACCATCCCATGATGGCCAGCATTTCT 41046
QY 425 att---ggcgttacgagccgatgggggtgtgtgctggtattccagctggagcgcgactt 481
DB 41045 GCTTCAACCGGCATGAGCCCGTTGGTGTCTGTGGCCAGATCATCCCGTGAACCTTCCCT 40986
QY 482 ttctttacgtcggtgggaagatagcccccgccctgcgcgcgcgtcctctcttcaatttca 541
DB 40985 TGGTCATGACGGTTGGAAACTTGGCCCGGCACCTGCCACAGGCAACACTTGGTTATGA 40926
QY 542 aagcctcggagaaaatcccccgtggcggttctgtggcgctcctcctctcttcgcagagccg 601
DB 40925 AGGTGGCAGACGACACCCCTCTCTGCCCTGATTTGGCTTCCCTCATCAAGAGGCAG 40866
QY 602 gattccctctggagctgagtcactcactcgtg---agcagagtgagcgggtgagcgt 658
DB 40865 GCTTTCCCTCGGTGGTGAACATCATACGGGTATGCCCCATGAGCAGAGTGGCGCCA 40806
QY 659 tggctcgacattggacattgcgaagatcagcttcacaaagatctgtcgscggtggcgcg 718
DB 40805 TCGCCACGACCGTGGATGTTGACAAAGTTGCTTCCCGGTTCCACCGAGTGGCGCAC 40746
QY 719 ccgtcaagcaagaacactcaagtccaacatgaagcgcgtcactctagaactgggggaaa 778
DB 40745 TGATCAGAAAGACGTGGCGATTCCAACTCAAGAGAGTACCTCGGTGGGTGGTGA 40686
QY 779 agccaacctcgtcttcaacgaagctcctctcgaacgcgcgtcgggggaatcggcaaggg 838
DB 40685 AGAGCCCCAGCATCGTGTGGCCGATGTGACATGGAGCATGCCGTGGAGCAGTCCACG 40626
```

```
QY 839 atttctcaaaattcgggcaaaatttggtccccccctccctggtttgtagtgcgaatgggaa 988
Db 40625 AAGCCCTGTTCTTC-----ACATGGGCGAGTGTGCTGTGCTGCCGACCTTCG 40572
QY 899 atttagcgagaaattccattgggtgcgtcgtcatttgagagctgtcagagatggc 958
Db 40571 TGAAGAAATCCATCTCAATAGTCTTCGAGAGAACCGGTGGAGAAAGCAAGCAGAGGA 40512
QY 959 ttggccagaaccattggaaccaagagagacgcatggtcccttcgtcgacaagtccca 1018
Db 40511 AAGTGGGGAACCCCTTTGAGCTGGACACCCAGCAGGGGCTCAGGTGGACAAAGGACAGT 40452
QY 1019 acacagagttctgggttaacattgacgttggc---aagataccgcgcagctcctcactg 1075
Db 40451 TTCAACGAGTCCTAGCTACATCCAGTGGCCAGAGAGGCGCAAAACTCTCTGTG 40392
QY 1076 gcttgtagaagggcgacaagggattcgattgaacgcagcatatttgcataatccaa 1135
Db 40391 GCGGAGAGCGTTTCGGGGAGCGGTGTTTTCATCAAGCCTACTGTTGTTGGTGGCGTGC 40332
QY 1136 aaccaggcagcaaaatttggtttgagagatctttggcccgctcttgcattaaagacgt 1195
Db 40331 AGGATGACATGAGAAATGCCAAAGAGGAGATCTTTGGGCTGTGACGCCCTTCAAGT 40272
QY 1196 tcaagcaggaagagggccattgagattgccaatgacacgaacttatgggctagcctcg 1255
Db 40271 TCAAGAAAGATTGAGAGGTGGTTGAGAGGGGCCAACCAACAGGATATGGCTGGCTGGCG 40212
QY 1256 teatttataccaaatctcaacaggggtctcgtgctcgtcgcgctcgagacgggtg 1315
Db 40211 CTGTGTTACCCGGGATCTGGACAAAGGCCATGTACTACCCAGGACATCCAGGCCGGA 40152
QY 1316 gctctcgatcaactccccctttatccccgagacacaaactcgtttggcgcatgaaac 1375
Db 40151 CCGTGGGGTAAACCTCAACATCGTCACTGCCACACCGCCATTTGGAGGGTTTAAAG 40092
QY 1376 aatcggtcagcgagagactagcgagaagggctcaagcgctacttggagcccaaga 1435
Db 40091 AATCTGGAACGGGAGGAGCTGGGTGAGATGGGGCTTAAGGCCTACACAGAGGTAAGA 40032
QY 1436 ccaataatccacgt 1451
Db 40031 CGGTACCATCAAGT 40016

RESULT 9
A43275 LOCUS
DEFINITION Sequence 1 from Patent WO9506122.
ACCESSION A43275
VERSION A43275.1 GI:2298573
KEYWORDS
SOURCE Alternaria alternata.
ORGANISM Alternaria alternata
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
Pleiosporales; Pleiosporaceae; mitosporic Pleiosporaceae; Alternaria.
REFERENCE 1 (bases 1 to 1485)
AUTHORS Achatz,G., Oberkofler,H., Simon,B., Unger,A., Lechenauer,F.,
Hirschwehr,R., Ebner,C., Kraft,D., Prillinger,H. and Breitenbach,M.
TITLE RECOMBINANT ALTERNARIA ALTERNATA ALLERGENES
JOURNAL Patent: WO 9506122-A 1 02-MAR-1995;
COMMENT BIOMAY PROD & HANDEL (AT)
Other publication FI 960882 960424
Other publication NO 960748 960424
Other publication AT 400723 960325
Other publication AT 172693 950715
Other publication AU 7377294 950321.
FEATURES
source
1..1485
/organism="Alternaria alternata"
/db_xref="taxon:5599"
/dev_stage="spores and vegetative hyphen"
BASE COUNT 352 a 423 c 386 g 324 t
```

ORIGIN

```
Query Match 16.3%; Score 239.2; DB 6; Length 1485;
Best Local Similarity 51.5%; Pred. No. 5e-48;
Matches 744; Conservative 0; Mismatches 663; Indels 37; Gaps 7;

QY 18 cgaatacaagagtgaaactcttccatcaacaatgaattcgtctctcccaagggttccgagag 77
Db 42 CGAGCAGCCACCGGACTCTTCAACAATAGTGTAAAGGCTGTTCACGGGCAAGAC 101
QY 78 attaaagctcaagaaacccgtggagcaaatccaccgttgcactgatttcaacgtggccaa 137
Db 102 CTTTGATGTTATCAACCCCTCCACTCAGGAGGTCATCTGCA---GTGTTCAGGAGGCCAC 158
QY 138 cggggccgattgtcgacagtgcagtagccgcttggtcgaggcggttcaaaaagggcccatg 197
Db 159 CGAGAAGGATGTGACATTTGCTTGTCTGCCGCCCAAGCCTTCAAC---GGCCCATG 215
QY 198 gaagaagttcaacaggtgcacaacgcgcggtgcagcttaagttcgcggaacctccga 257
Db 216 GGCARAAGGAGACACCAGACAGAGGGAAGCTGTAAACAGCTTGGCGATCTGTCCA 275
QY 258 gaagaacgcggaagactcgtcgtctggagtcgtgcgccacgcggtagaccggttgcgat 317
Db 276 GAAGAATGCGGACCTCATTTGCTGTGAGAGCTCTCGACACGCAAGGCCCTTCAGCAT 335
QY 318 gatcactcttccgacattccaaacgatgtccggttctcgtactactacagagctgggc 377
Db 336 GGCARAAGAGTGTGATTTCCCGCCCGCTGGTTGCCCTGAGGTACTACGGAGATGGGC 395
QY 378 cgacaagatcgccggagaagacctttcccgagagacaaacggaagccgaattggcgttacga 437
Db 396 CGACAAGATTGAGGCAAGGTGTCGACACAGCACAGCACCCGACAGCTTCAACTACATCCGCA 455
QY 438 gccgatgggggtgtgctgggtatttcgacgtggaacgcgaactttcttcaagtcggcgtg 497
Db 456 GAGCCTATTGGTGTGTTTGGGTGATCATCCATGGAACCTTCCCTATTCTTCATGFGTCAAG 515
QY 498 gaagatagccccgcgcctgcgcgcggtgctcctctcatcttcaagcctcgagagaatc 557
Db 516 GAAGATTGCTCTGCCATCGCCACTGCCACTAACACACCGTCTGCTCAAGACTGCTGAGCAGAC 575
QY 558 cccgtggggttcttggtggtcgtcctctcttcgcagaagcgcggaattccctccttgagt 617
Db 576 ACCTCTCTCCGCATACATTGCCCTGCAAGTGTATCCAGGAGGCGGTTTCCACACAGGTGT 635
QY 618 cgtgcagtttccctcaactggagcacagtgta---cgggtgaagcatttggctcgacatgga 674
Db 636 CATCAACGTCATCAGTGGTTTCGGAAGATCGCCGCTGCTGCCATGTCCGCTCAGATGGA 695
QY 675 cattcggaagatcagttcacaagatctctgcgggtggtggccgcgcgtcaagaagaacac 734
Db 696 CATTCACAAGATTGCTTTACTGGTTCAACCGCTTGTCCGCCGCTCAATCATGAAGTCTGC 755
QY 735 actcaagttccaatgaacgcgcgtcactatgaactgggggaaagcccaacca-----t 788
Db 756 CGTGGCTCCAACCTTGAAGAAGTCACTCTTGAGCTCGGAGGCAAGAGCCCAACATTTGT 815
QY 789 cgtcttcaacgaagctcctctcgaacgcgcagtcggtgggggaatcggcgaagatttccaa 848
Db 816 CTTTCGCGCAGCGAGATCTTTCAGCAGGCTATCCACTGGGTCAACTTTGGTATTACTTCAA 875
QY 849 attcgggcaaatgggtcccccctcctcttctgtagtgcattggggaattatgacga 908
Db 876 CCACGGACAGGCTTGTGTGCTGGTTCGGGTATCTACAGTCCCAAGAGAGATCTACACAAA 935
QY 909 gaaatccatggagtc--cgtcatggctcatttggaggtcgtcagagatggcttggccag 966
Db 936 GTTATCCAGCGCTTCAAGGAGCGGCTGCTCAGAACCGCTTGTGGTGAACCCATTCGCCGC 995
QY 967 aaccatgtgaacccaagagagcagcatgctcctctcgtcgaacagtcaccagtacgaca 1026
```



```
QY 849 attcgggcaaatbttgggtccccccctcctgttctgttagtgcgaatggggaaatttagcgga 908
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 934 CCACGACAGAGCGCTTGTGTGCTGGTTCGCTATCTACGTCCCAAGAAGAGATCTACGACAA 993
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 909 gaaattccattgagtc--cgtaagcctcatttgaggagctcagagatgagctggccag 966
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 994 GTTCATCAGCGCTTCAAGAGACGGGCTGCTCAGAACCGCTTGTGTGACCCATTCGCGCG 1053
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 967 aaccattgggaacccaagagacgcatggtcccttcgtcgcgaagtcccaagtaagacaga 1026
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1054 GACAC-----TCCAGGGTCTCAAGTCTCGCAGCTCCAGTTCGAGCGT 1096
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1027 gtcttggttaacattgaattgggaagga---taccgagagctcctcactggcggttgg 1083
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1097 ATCATGGGCTACATCGAGGAGGGCAAGATCTGGCGCGACCATCGAGACTGGTGGCAAC 1156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1084 agaaaggcgacaaggattgcgattgaaccgagcatattgtccaatcccaaaccaaggc 1143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1157 CGTAAGGCTGACAGGGTTACTTCATCGAGCCACAACTTCTCCACGTAACCGAGGAC 1216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1144 agcaaaatttggttagagagatotttgccccctgttccattgaagacgttcaagacg 1203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1217 ATGAAGATTCAAGCAAGAGAGATCTCGGCCCGCTCTGCACAATCTCCAAGTTCAAGACA 1276
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1204 gaagaaggagccattgagattgccaatgacacgactlatggctagctcggtcatattat 1263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1277 AAGCGGACGTCATCAAGATTGGCAACACACCATACATAGGTCCTTTCGGCCGCTGTACAC 1336
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1264 accaaattctcaacagggttcgtgtctcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1323
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1337 ACATCCAAACCTCACACACTGCCACTGCCATGAAGTTGCCAAGCGGCTCCGTGCGAGGAATGTCGTGG 1396
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1324 atcaacttccccctttatccccgagacacaaactccgttttgccgagcatgaacaacatcggc 1383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1397 GTCAACTCTTACACACTTTCACATGCGCAGCTTCCCTTCGGAGGGTACAAAGAGTCTGGT 1456
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1384 tcaggcagagctagggcgaagggtcctcaaggcgtacttggagcccaagaccattaat 1443
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1457 ATTGGCGCGAGTTGGGAGAGCGCGCTGGACAACATACATCCAGACCAAGACCGGTGCT 1516
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1444 atcc 1447
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1517 ATTC 1520
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 11

HUMTALD 6074 bp DNA linear PRI 15-APR-1996
LOCUS Human mitochondrial aldehyde dehydrogenase x gene, complete cds.

DEFINITION M63967
ACCESSION M63967
VERSION 1
KEYWORDS aldehyde dehydrogenase.
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 6074)

REFERENCE Hsu,L.C. and Chang,W.C.
AUTHORS Cloning and characterization of a new functional human aldehyde
TITLE dehydrogenase gene
J. Biol. Chem. 266 (19), 12257-12265 (1991)

FEATURES 91286241 Location/Qualifiers

1..6074

/organism="Homo sapiens"

/db_xref="taxon:9606"

1301..1317

2257..3966

2266..3819

/codon_start=1

/product="aldehyde dehydrogenase"

/protein_id="AAA96830.1"

/db_xref="GI:1263008"

source

exon

exon

CDS

/translation="MLRFLAPRLSLQRTALYSAAALFSPILNPDIPTYNOLFINNE
WQAVSKTPTPTNPTTGEVIGHVAGDRAVDRAVKAAREAFRLGSPRRMDASERG
RLNLLADLVERDRVYLASLETLDNGKPFQESALDLDDEVIKVRYFAGWADKHGKT
IPMHGQHFCTRHEPVCVCGOIIPWNPFLMVGWKLAPALATGNTVVKVAEOTPLSA
LYLASLIKEAGFPFGVYNIITGPTAGAAIAAHMDVDKVAFTGSTIEVSHLIQKAAGD
SNLRKVTLELGGKSPSIVLADADMEHAHEVQECHEALFVNMGCCGCCAGSRTFVEESIYNE
FLBERTKAKORKVGNFELDTQGGPOVDKEQFPRVLGYIOLGQKGAKLCCGGERFG
ERGFIFKPTFGVQDDIMRIAKEIFGPVQPLFKFKIEEVVERANTRYGLAAAVFT
RDLDKAMFTQALQAGTVMVNTYNIIVTCHTPFGGFKESGNGRELGEDGLKAYTEVKT
TIKVPKNS"
variation 2448
variation 2552
variation 2585
repeat_region 4803..5137
repeat_region 5613..5924
BASE COUNT 1475 a 1505 c 1594 g 1500 t
ORIGIN

Query Match 16.3%; Score 238.8; DB 9; Length 6074;
Best Local Similarity 49.7%; Pred. No. 6.7e-48;
Matches 724; Conservative 0; Mismatches 717; Indels 15; Gaps 4;

QY 5 ttctttgcctgacgaatcacagagtgaaactctctcatcaacaatgaattcgtctctcca 64
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2351 TTCTGAACCCAGACATCCCTTACACACAGCTGTTTCATCACAATGAATGGCAAGATGCAG 2410

QY 65 aggggtccgagatgaaacgctcacgaaccggtggagcaatccacggttccactgatg 124
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2411 TCAGCAAGAAGACCTTCCGACGGTCAACCTACCACCGGGGAGGTCTATCGGGCACGTGG 2470

QY 125 tteacgtggcacaacgcgccgatctgcagatcagtagcagcttcggtcgagcggtca 184
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2471 CTGAAGTGTGACCGGCTCATGTGGATCGGGCGGTGAAGACGCCCGGAAAGCGCTTCCGCC 2530

QY 185 aaaagggtccctggaagaagtctcacaggtgcacaacgcgcgcgtgcattgcttaagtctg 244
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2531 TGGGTCCCCATGGCGCGGATGGATGCTCTGAGCGGGGCGGCTGCTGAACCTCCTGG 2590

QY 245 cggacctcgcgagaagaacccgagagctcgtcgtctggagtcgctgccaccgcta 304
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2591 CAGACCTAGTGGAGCGGATCAGTCTACTTTGGGCTCTACTCGAGACCTTGGACAATGGGA 2650

QY 305 gaccggtgtcgtatgatcaactcatcttcacattccaaacatggtctcgtgttctcgtact 364
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2651 AGCCTTTCAGAGAGTCTTACGCCTTGGACTTGGATGAGGTCTCATCAAGGTGTATCGGTACT 2710

QY 365 atgcaggtggccgacagaatgcgcggaaagaccttccccgagagacaacgcgcgcga 424
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2711 TTGCTGGTGGGCTGACAAGTGGCATGGCAAGACCATCCCATGTCATGGCCAGCATTTTCT 2770

QY 425 att--ggcgttacgagccgatgggggtgtgtgtgcttggtatgccagctgggaacgcgactt 481
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2771 GCTTACCCCGCATGAGCCCGTGGTGTGTGTGGCCAGATCATCCCGTGGAACTTCCCT 2830

QY 482 ttcttaagtcggtggaagatagccccgccttcgcgcgcgcgcgcgcgcgcgcgcgcgc 541
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2831 TGGTCATGCAGGGTTGGAACACTTGCCCGGCACATCGCCACAGGCAACACTGTGGTTATGA 2890

QY 542 aagcctcgagaatccccgcgtggcggtcttcggcctgcctcctctcttcgcaaaagccg 601
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2891 AGGTGGCAGACACACCCCTCTCTGTGCCCTGTATTTGGCTCTCCCTCATCAAGGAGGCGAG 2950

QY 602 gattccctctcgtgagtcgtgcagttctcactgg---agcacgagtgacgggtgaagcat 658
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2951 GCTTTCCCTCTGGGTGTGAACATCATCACGGGTATGGCCCAACACAGAGGTGGCGCCA 3010

QY 659 tggcgtcgcacattggacaattgcgaagatcgaagatctgtcgcgcggtggcgcg 718
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3011 TCGCCACGACATGATGTGTGACAAGTTGCTTTACCGGTTCCACCGAGGTGGGCCACC 3070

QY 719 ccgtcaagcaagcaacactcaagtcacaacatgaagcgctcactctagaactggggaaa 778
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

| | | | |
|--------|--|------|--|
| gene | 5791..7688 | gene | /number=1 |
| exon | /gene="B11B23.040" | CDS | /complement(11719..12591) |
| gene | /complement(5791..6147) | | /gene="B11B23.060" |
| CDS | /gene="B11B23.040" | | /complement(11719..12591) |
| | /number=1 | | /gene="B11B23.060" |
| | complement(join(5791..6147,6331..6937,7024..7688)) | | /codon_start=1 |
| | /gene="B11B23.040" | | /product="hypothetical protein" |
| | complement(join(5791..6147,6331..6937,7024..7688)) | | /protein_id="CAD21123.1" |
| | /gene="B11B23.040" | | /db_xref="GI:18376345" |
| | /note="similarity to chromosome 5, pl clone:mx110, arabidopsis thaliana, TREMBL:ATAB5248_2" | gene | /translation="MATPTPTTNNKGDIDPTSLAKTIAMKHKATDOLLHMLIHS |
| | /codon_start=1 | gene | PLALLQLDLLEKPVRLQVLDLGEDVLRVDMSLAEKLVQTRAKGAFQTRTTIAH |
| | /protein_id="CAD21121.1" | gene | TIPIHPMKSWAERERLVAMTKAREGLGLVGEVGAEEGKKKEQKVKREKGG |
| | /db_xref="GI:18376343" | exon | GGAVIDTYLAPLFDWDGAVSAISVAFFGCVAAKGGVVEQGLMAWETQKMWVPGGAD |
| | /translation="MSPLNLNLNLTYQNIHFHPSSSSSDPIPLPPGISRSFI | CDS | GFKVLVNAWFPVLGIHIAEMVNEPRTLRANKDLKMLGLTLWLMWVVFEGYMAFR |
| intron | PTPHGDEIHTATTTTSSSQPALFFIHGGMGASVWLEYLFFSQQQQQQNI | | RYDEKAAELERKQH |
| exon | PCYAIWRHGGSWYPSYLRMYVTGKGLDGLDVGVSGIRWAVDRGQRKQLQQLGRE | | complement(13073..14179) |
| intron | GIKQGGEGGGEWEVEMVLIGHSGGGLAQYIISAGLGLGLGYKIKGLVLMGAVP | | /gene="B11B23.070" |
| exon | GFGSIPITLWFRIDPLFALRLLHLGHPNSPLSHPSLVRLLIFFSSNDLLPDSYLSRF | | 13073..14179 |
| exon | MSFMSRYSLLPLMGARSFADPKVLLRISGWEETTKQEQKQGOEERVLVIAGORD | | /gene="B11B23.070" |
| exon | KIMTQPTMORLADWYRAAYORLVVEKKIDGVVMVGSSENETVDEEVEVEVEGSGGF | | complement(13073..14179) |
| exon | KDTAGRGVREALVPGAGHCGNDGHEVGARKPVCLLNTLRKWCAGAGVAILRDEYG | | /gene="B11B23.070" |
| exon | VPMRSVGETSHKVIQGVIPVKVHLITDSLLQLTHFGISSHATSMPTITNILLISFVAS | | /codon_start=1 |
| exon | ALSFDCLUSTALSVAGIKPNDTERTERYTKPTGTQK | | /product="hypothetical protein" |
| exon | complement(6148..6330) | | /protein_id="CAD21122.1" |
| exon | /gene="B11B23.040" | | /db_xref="GI:18376346" |
| exon | /number=1 | | /translation="MVSAAEPSPFSESNKFCLYSHGPFKIIQRTQHIPIVMALPVST |
| exon | complement(6331..6937) | | VALLEIFVDANPEVWFSRLDOAHCDNLNRAKAAVHRHEIFLDLLOKHVKPQGHVLI |
| exon | /gene="B11B23.040" | | SAQTHPSFPRVLARAMYKDKSAIMCLPHCSDTILSKEEKEVDLTCKRCRAQVETMSEY |
| exon | /number=2 | | DMWRMHHEEGVLEENGRAPIVTPGPEWOSLIREEDVAISAMIVETKICORQEVRR |
| exon | complement(6938..7023) | | RAKISPTGLPWEELKEETEEIRKLLKLLGSLNLYLNSRNKSQSLKVLMDIDPVR |
| exon | /gene="B11B23.040" | | AYLSATVHLEWATANLQLASSSCLLEIPKPNKARKVSDGGGILRRYVAALVCAEKGSQ |
| exon | /number=2 | | |
| exon | complement(7024..7688) | | |
| exon | /gene="B11B23.040" | | |
| exon | /number=3 | | |
| exon | 8798..8857 | | |
| exon | /gene="B11B23.050" | | |
| exon | /number=1 | | |
| exon | 8798..10714 | | |
| exon | /gene="B11B23.050" | | |
| exon | join(8798..8857,8921..10714) | | |
| exon | /gene="B11B23.050" | | |
| exon | /note="similarity to lariat-debranching enzyme, YKL149c, Saccharomyces cerevisiae, PIR:A38491" | | |
| exon | /codon_start=1 | | |
| exon | /product="related to lariat-debranching enzyme" | | |
| exon | /protein_id="CAD21122.1" | | |
| exon | /db_xref="GI:18376344" | | |
| exon | /translation="MSDIDITQOQVRIAVEGCHGHTLNAIYAATEKSAKERGWGV | | |
| exon | DLIIIGDFOAVRNAADLNAMSPVKYRELGDFEYYSGVKRAPIYLTIFVAGNHFAAS | | |
| exon | KLWELYGWVCPNLIYMGANVLRGLPLRAGMSGIWKGYNYRKAHERLFPNDQDV | | |
| exon | KSFTHREIDVRKELQIQVDIGISHDWPRATEKHGNTKHLNMRKPFERESADGSL | | |
| exon | GQAAEYMDLRPPYWFSAHLHCKYSAIKTYDLPKETTSDGGETTKPEETAVPQOE | | |
| exon | QOAPAAANDEIDLDMEDFAPAPQOATAPVATSNPDEIDLDDDEVAPTAPSQAPT | | |
| exon | SAAFAAVTEDEPTALRSLIPASFPNPNQSYNAQPHGNSGNKPKPGVPPTI | | |
| exon | TNTVVRFLSKDKUCPLRQLAEIEPINDSPKELQRLSLSYDPEWLSITLAFHPLL | | |
| exon | RSVAGKPCSPHNAPIPVDLGEAGYKPIIEARKWVDEHITGDKLLVPDFVOTAPPI | | |
| exon | DLVNRPEGNMTDEMPYETNPHTAEFCRLIGVENVYNDASEERGRVRGPRPMAER | | |
| exon | EGGRCGHRGGGGGGGGGRRGGRGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG | | |
| exon | 8858..8920 | | |
| exon | /gene="B11B23.050" | | |
| exon | /number=1 | | |
| exon | 8921..10714 | | |
| exon | /gene="B11B23.050" | | |
| exon | /number=2 | | |
| exon | 10924..11043 | | |
| exon | /product="5S ribosomal RNA" | | |
| exon | 11719..12591 | | |
| exon | /gene="B11B23.060" | | |
| exon | complement(11719..12591) | | |
| exon | /gene="B11B23.060" | | |

| | | | | | | | |
|-----------------------|-------|--|---------|------------|-----|--------|-------|
| Query Match | 16.3% | Score | 238.2 | DB | 8 | Length | 79554 |
| Best Local Similarity | 53.3% | Pred. No. | 1.1e-47 | | | | |
| Matches | 694 | Conservative | 0 | Mismatches | 583 | Indels | 26 |
| Gaps | 8 | | | | | | |
| QY | 37 | ttoatcaacaatgaattogtctctcccaagggtccgagagattacacgtcacgaaccg | 96 | | | | |
| DB | 25397 | TTTCATAACACAGAGTTCGTGAGGGGTGTCGACAGAGACCTTCGAGGTTCATCAACCC | 25338 | | | | |
| QY | 97 | tgggacgaatccacgttcacgttcacgttcacgttcacgttcacgttcacgttcacgt | 156 | | | | |
| DB | 25337 | GCCACGAGGAGTCACTGCTC---CGTCCACGAGGCCACGAGAGGATGCGATATC | 25281 | | | | |
| QY | 157 | gcagtagccgcttcggtgcagcggtccaaagggcccatggaagaagtccacaggtgca | 216 | | | | |
| DB | 25280 | GCTGTCGCGCTCCCGCAAGGCTTTC---GAGGGTGTCTGGCGCGAGCTCCTCCCGAG | 25224 | | | | |
| QY | 217 | caacgcgcgcgtgcgttcggttcggttcggttcggttcggttcggttcggttcggttcg | 276 | | | | |
| DB | 25223 | CACGCTGGCATCTACCTCTCTCAAGTCGCGGACCTGCTCGAGAGAACTCGACTCCTT | 25164 | | | | |
| QY | 277 | gtcgtctggagtcgctgc | 336 | | | | |
| DB | 25163 | GCGGCTGTCAGTCTCTGCACAAATGGCAAGTCCATCACCATTG---GCCGCTGGCATGT | 25107 | | | | |
| QY | 337 | caaaatgcttcggttcggttcggttcggttcggttcggttcggttcggttcggttcggttc | 396 | | | | |
| DB | 25106 | GCGGCGCTGCTGGCACTATCCGCTACTACGCGGCTGGCGCGACAGATCGAGGGCAAG | 25047 | | | | |
| QY | 397 | acctt---tcccgagagacaacgcggaatggttcagccggttcagccggttcggttcggttc | 453 | | | | |
| DB | 25046 | ACCATTGATATCTCCCGGACCTCTTCCACTACACGAGCAAGACCTCTCGGTCTTTGC | 24987 | | | | |
| QY | 454 | gtcgttcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtc | 513 | | | | |
| DB | 24986 | GCTCAGATCATCCCTGGAACCTTCCTCTCCCTCATGCTTGCTGGAAGGTGGGGCCCC | 24927 | | | | |
| QY | 514 | ctgc | 573 | | | | |

[illegible]

USA
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgapbs@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyak, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK plate: 37 Row: 9 Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: Similarity but not
identity to protein.

```

FEATURES             source
location/qualifiers
1..2236
/organism="Mus musculus"
/db_xref="LocusID:72535"
/db_xref="taxon:10090"
/clone="MGC:28561 IMAGE:4207479"
/tissue_type="Colon, normal, 5 month old male mouse."
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
62..1621
/codon_start=1
/product="aldehyde dehydrogenase 1 family, member B1"
/protein_id="AAH20001.1"
/db_xref="GI:18043461"
/translation="MLTARLLPLRLLCQGRTTSTSTAAALPNPILNPETCYNKLFIN
NGWHDASKKTPTVNTPTGEIVGHVAGEDRADVDLVAARAAEAFPLGSPWRMRDASE
RGLLRNLADLVVERDVTYLAETLNDGKPFQESYVLDDDEVIKVYFAGWADKWHG
KITPMGEHFCFTRHEPQVCGCQIIPMNEPLVMQGWKLAPALATGNTVYKVAEJOPLA
SALYLASLTLEKAGFPQGVNITGVCPTAGAAIAOHMDVDKYVNTGSTVGHJLOKPA
GSNLRATVLELGGKSPSYLLADAMEHAVDOCHEALFNNMGQCCACGSRTPVEESII
RFLERFTEBKAKQKRVGDFELDTQQGPDVQKPEFRIIGYILNGKEAGKLLCGGER
LGERGFEFIKPTVFGDQMGKAKEEIPGVQPLFKFKKIEEYIQRRNNTYRGLAAAY
FTRDLKATYFTQALQAGVWVNTYINVTCTPTFGGFKESGNGRELGDGLRAYTEVET
TWTIKYPEKNS"
562 a 571 c 644 g 459 t
BASE COUNT
ORIGIN

```

| Query Match | 15.8% | Score 232: | DB 10: | Length 2236: |
|--|-------|------------------|------------|--------------|
| Best Local Similarity | 51.0% | Pred. No. 3e-46: | | |
| Matches 684: Conservative | 0: | Mismatches 640: | Indels 18: | Gaps 5: |
| QY 122 atgtctacgtggccaaacgcccgcgatgtctgcacagtgcagtagccgcttcggtgcagcgcg 181 | | | | |
| Db 267 ATGTGGCGAAGGTGACCGGCGACATGTGGATCTGGCTGTAAAGACAGCCGAGACGCT 326 | | | | |
| QY 182 tcaaaaagg - -cccatgaaagaagttccacagtgccacacgcgcgcgcgtgcattgta 238 | | | | |
| Db 327 TCCGCTGGGGTCCCCATGCGCAGGATGGATGCCTCAGAGCGGGCCCGGCTCGTGAACC 386 | | | | |
| QY 239 agltcgcggacctgcgcagaagaacgcgcgaagctcgcctctctgcagtcgctgcacca 298 | | | | |
| Db 387 GCCTAGCTGATCTTTGGGAACGACATCCGAGTGTACTTGGCCTCCTCTGAGACCGCTAGATA 446 | | | | |
| QY 299 ccggttagacggtgtcgatgatcactcatttcgacattccaaacaatggctcccgctgttc 358 | | | | |
| Db 447 ACGGGAACCTTTCCAGGAGCTTTATGTCTTGGATCTGGATGAAGTCATCAAGGTGTACC 506 | | | | |
| QY 359 gctactatgcaggctgggcgcgaacatgcgcggaaagacctttcccgaggacaacaggca 418 | | | | |

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 02:58:36 ; Search time 4999.95 Seconds
(without alignments)
3951.948 Million cell updates/sec

Title: US-09-882-694A-4
Perfect score: 1464
Sequence: 1 atggtcttcctgcctgaacga.....tccacgtcaacatagatga 1464

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:**
 - 2: em_esthum:**
 - 3: em_estin:**
 - 4: em_estmu:**
 - 5: em_estov:**
 - 6: em_estpl:**
 - 7: em_estro:**
 - 8: em_htc:**
 - 9: gb_estl:**
 - 10: gb_est2:**
 - 11: gb_htc:**
 - 12: gb_gss:**
 - 13: em_gss_hum:**
 - 14: em_gss_inv:**
 - 15: em_gss_pin:**
 - 16: em_gss_vrt:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------|--------------------|
| 1 | 22 | 1.5 | 566 | 9 AU204674 | AU204674 AU204674 |
| 2 | 21 | 1.4 | 378 | 12 AQ848277 | AQ848277 LMAJFV1_1 |
| 3 | 21 | 1.4 | 524 | 10 BG557867 | BG557867 EML_55_E1 |
| 4 | 21 | 1.4 | 757 | 9 AI068883 | AI068883 mgae0004c |
| 5 | 21 | 1.4 | 769 | 9 BI225854 | BI225854 602950228 |
| 6 | 21 | 1.4 | 775 | 10 BI856603 | BI856603 603385735 |
| 7 | 21 | 1.4 | 801 | 9 AI068982 | AI068982 mgae0004d |
| 8 | 21 | 1.4 | 852 | 10 BI258465 | BI258465 602972456 |
| 9 | 21 | 1.4 | 930 | 10 BF035820 | BF035820 601458344 |
| 10 | 21 | 1.4 | 938 | 9 AL536113 | AL536113 AL536113 |
| 11 | 21 | 1.4 | 1243 | 10 BI457385 | BI457385 603185515 |
| 12 | 20 | 1.4 | 534 | 10 BG816128 | BG816128 dac07h02. |
| 13 | 20 | 1.4 | 572 | 9 AI293860 | AI293860 LP07108.3 |
| 14 | 20 | 1.4 | 582 | 10 BJ060784 | BJ060784 BJ060784 |
| 15 | 20 | 1.4 | 583 | 9 AI259976 | AI259976 LP03541.5 |
| 16 | 20 | 1.4 | 583 | 10 BI942019 | BI942019 se82c09.y |
| 17 | 20 | 1.4 | 591 | 10 BG905010 | BG905010 tatrl136E |

| | | | | | |
|----|----|-----|-----|-------------|--------------------|
| 18 | 20 | 1.4 | 595 | 9 AI261062 | AI261062 LP05133.5 |
| 19 | 20 | 1.4 | 597 | 9 AI114304 | AI114304 GH11111.3 |
| 20 | 20 | 1.4 | 604 | 9 AI296186 | AI296186 LP10021.5 |
| 21 | 20 | 1.4 | 630 | 9 AI455414 | AI455414 LD23985.3 |
| 22 | 20 | 1.4 | 677 | 10 BE282264 | BE282264 601101239 |
| 23 | 19 | 1.3 | 115 | 10 BJ112053 | BJ112053 BJ112053 |
| 24 | 19 | 1.3 | 188 | 9 AW230296 | AW230296 up28g10.y |
| 25 | 19 | 1.3 | 201 | 10 BJ121003 | BJ121003 BJ121003 |
| 26 | 19 | 1.3 | 203 | 9 AV248967 | AV248967 AV248967 |
| 27 | 19 | 1.3 | 207 | 10 BE439065 | BE439065 CDO269.WH |
| 28 | 19 | 1.3 | 209 | 10 BM090714 | BM090714 Iq17h05.y |
| 29 | 19 | 1.3 | 218 | 9 BB117397 | BB117397 BB117397 |
| 30 | 19 | 1.3 | 237 | 9 BB565455 | BB565455 BB565455 |
| 31 | 19 | 1.3 | 252 | 10 BJ101954 | BJ101954 BJ101954 |
| 32 | 19 | 1.3 | 261 | 9 AA197346 | AA197346 zp92f10.s |
| 33 | 19 | 1.3 | 308 | 10 M75758 | M75758 CEL01A951.C |
| 34 | 19 | 1.3 | 310 | 9 BB377860 | BB377860 BB377860 |
| 35 | 19 | 1.3 | 358 | 9 AU109836 | AU109836 AU109836 |
| 36 | 19 | 1.3 | 360 | 10 C49496 | C49496 C49496 yuji |
| 37 | 19 | 1.3 | 361 | 9 AU109727 | AU109727 AU109727 |
| 38 | 19 | 1.3 | 367 | 9 AW897385 | AW897385 RC4-NN005 |
| 39 | 19 | 1.3 | 369 | 10 BE439116 | BE439116 CDO269.WH |
| 40 | 19 | 1.3 | 388 | 10 H79241 | H79241 yu27g08.r1 |
| 41 | 19 | 1.3 | 401 | 9 AW847618 | AW847618 IL3-CT021 |
| 42 | 19 | 1.3 | 401 | 9 AW847619 | AW847619 IL3-CT021 |
| 43 | 19 | 1.3 | 401 | 9 AW847629 | AW847629 IL3-CT021 |
| 44 | 19 | 1.3 | 401 | 9 AW847634 | AW847634 IL3-CT021 |
| 45 | 19 | 1.3 | 413 | 9 AW847540 | AW847540 IL3-CT021 |

ALIGNMENTS

RESULT 1

AU204674
LOCUS AU204674 566 bp mRNA linear EST 17-JUL-2001
DEFINITION AU204674 unpublished oligo-capped cDNA library, stage L4
Caenorhabditis elegans cDNA clone yk840d11 5', mRNA sequence.
ACCESSION AU204674
VERSION AU204674.1 GI:14836273
SOURCE EST.
ORGANISM Caenorhabditis elegans.
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
; Rhabditidae; Pelodierinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 566)
AUTHORS Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE A complementary view of the C.elegans genome
JOURNAL Unpublished (2001)
COMMENT Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES

source
1..566
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk840d11"
/clone_lib="unpublished oligo-capped cDNA library, stage L4"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L4"
BASE COUNT 160 a 129 g 148 t
ORIGIN
1.5% Score 22; DB 9; Length 566;

Query Match

```

Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 35 tcttcatacaaatgaattcgt 56
|||||
Db 118 TCTTCATCAACAATGAATTCGT 139

RESULT 2
AQ848277/c
LOCUS AQ848277 378 bp DNA linear GSS 25-MAY-2001
DEFINITION LMAJFV1_lm59a07.x1 Leishmania major FV1 random genomic library
Leishmania major genomic clone LMAJFV1_lm59a07 3' similar to
SW:DHAM.LEITA Q25417 ALDEHYDE DEHYDROGENASE, MITOCHONDRIAL
PRECUSOR ;, DNA sequence.
ACCESSION AQ848277
VERSION AQ848277.1 GI:5052886
KEYWORDS GSS.
SOURCE Leishmania major.
ORGANISM Leishmania major
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE 1 (bases 1 to 378)
AUTHORS Akopyants,N.S., Clifton,S.W., Martin,J., Pape,D., Wylie,T., Li,L.,
Kissinger,J., Roos,D.S., Marra,M., Hillier,L., Chinwalla,A.,
Bilstain,A., Schmitt,A., Person,B., Theising,B., Ritter,E., Ronko
i., Bennett,J., Cole,R., Underwood,K., Cardenas,M., Gibbons,M.,
Harvey,N., McCann,R., Tsagarishvili,R., Williams,T., Jackson,Y.,
Bowers,Y., Swaller,T., Waterston,R., Wilson,R. and Beverley,S.M.,
A survey of the Leishmania major Friedlin strain V1 genome by
shotgun sequencing: a resource for DNA microarrays and expression
profiling
Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)
JOURNAL 21192569
MEDLINE
COMMENT Contact: Akopyants, NS / Beverley, SM
WashU Leishmania Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewartson.wustli.edu
Library construction: Natalia S. Akopyants, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
If using this information please cite:
N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major
Friedlin strain V1 genome by shotgun sequencing' and the Washington
University Genome Sequencing Center For information on obtaining
clone material please contact: Natalia S. Akopyants Ph.D.
(natalia@borcim.wustli.edu) and/or Stephen M. Beverley Ph.D.
(beverley@borcim.wustli.edu)
Seq primer: -40up from Gibco
Class: shotgun
High quality sequence stop: 82.
FEATURES
source
Location/Qualifiers
1..378
/organism="Leishmania major"
/strain="Friedlin strain V1"
/db_xref="taxon:5664"
/clone="LMAJFV1_lm59a07"
/clone_lib="LMAJFV1_lm59a07"
/lab_host="TOP10 (Invitrogen)"
/note="vector: pZero-2 (Invitrogen); Site_1: EcoRV;
Genomic DNA was isolated from stationary phase cells. For
this library, DNA was sheared to give a tight size
distribution of 1-1.5kb fragments, blunt-ended with T4 DNA
polymerase, dephosphorylated with Shrimp Alkaline
Phosphatase and ligated into pZero-2 vector's EcoRV site."
BASE COUNT 77 a 117 c 104 g 80 t
Query Match 1.4%; Score 21; DB 12; Length 378;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

us-09-882-694a-4.rst
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1159 gagagatctttggcccgctc 1179
|||||
Db 306 GAGGAGATCTTTGGCCCCGTC 286

RESULT 3
BG557867
LOCUS EMI_55_E11.b1_A002 Embryo 1 (EMI) Sorghum bicolor cDNA, mRNA
DEFINITION BG557867
ACCESSION BG557867
VERSION BG557867.1 GI:13586865
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 524)
AUTHORS Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
TITLE An EST database from Sorghum: developing embryos
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude polyA, vector and regions
below phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 491
POLYA=No.
FEATURES
source
Location/Qualifiers
1..524
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Embryo 1 (EMI)"
/note="Organ: Embryos germinated for 24 hr; Vector:
pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
EcoRI; The library was made from poly-A RNA in the cloning
vector Lambda ZAP II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT 68 a 206 c 155 g 95 t
Query Match 1.4%; Score 21; DB 10; Length 524;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 505 gcccccgcctcgccgcggc 525
|||||
Db 30 GCCCCCGCCTCGCCGCGGC 50

RESULT 4
AI068883
LOCUS AI068883 757 bp mRNA linear EST 09-DEC-1999
DEFINITION mgae0004cb07f Magnaporthe grisea Appressorium Stage cDNA Library
Magnaporthe grisea clone mgae0004cd07f 5', mRNA sequence.
ACCESSION AI068883
VERSION AI068883.1 GI:3391858
KEYWORDS EST.
SOURCE Magnaporthe grisea.
ORGANISM Magnaporthe grisea
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.
REFERENCE 1 (bases 1 to 757)

```

AUTHORS
TITLE
JOURNAL
COMMENT

Choi, W., Fang, E., Sasnowski, M., Wing, R. and Dean, R.A.
 Expressed sequence characterization during appressorium formation
 in rice blast fungus, Magnaporthe grisea
 Unpublished (1998)
 Contact: Dean, R.A.
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson University, Clemson, SC 29634, USA
 Tel: 864 656 5737
 Fax: 864 656 4293
 Email: rdean@clemson.edu
 Seq primer: T3 primer (AATTAACCTCTACTAAAGG)
 High quality sequence stop: 325.

FEATURES
 source

1..757
 /organism="Magnaporthe grisea"
 /strain="70-15"
 /db_xref="taxon:148305"
 /clone="mdae0004cd07f"
 /clone_lib="Magnaporthe grisea Appressorium Stage cDNA
 Library"
 /dev_stage="Germinated conidia on appressorium-inductive
 surface"
 /note="Vector: pBlueScriptII SK(+) Vector; Site_1: EcoRI;
 Site_2: XhoI; The appressorium formation-specific cDNA
 library was constructed from conidia germinated for 5-8
 hr on an inductive surface. The library has an average
 insert size of 1.5 kbp."
 BASE COUNT 181 a 218 c 196 g 161 t 1 others

Query Match
 Best Local Similarity 1.4%; Score 21; DB 9; Length 757;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1159 gaggagatcttggccgcgtc 1179
 |||||
 Db 431 GAGGAGATCTTTGGCCCCGTC 451

RESULT 5
LOCUS BI225854/c
DEFINITION 602950228F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:5094559 5',
 mRNA sequence.
ACCESSION BI225854
VERSION BI225854.1 GI:14679298
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 769)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: DCTD/DTF
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLNL1963 row: g column: 08
 High quality sequence stop: 723.

FEATURES
 source

1..769
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5094559"

/clone_lib="NIH_MGC_8"
 /tissue_type="Burkitt lymphoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5',
 adaptor: GGCAGGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 182 a 208 c 223 g 156 t

Query Match
 Best Local Similarity 1.4%; Score 21; DB 10; Length 769;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 895 ggaatttagcgaggaatttc 915
 |||||
 Db 168 GGAATTAGCGGAGAATTC 148

RESULT 6
LOCUS BI856603/c
DEFINITION 603385735F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5394441 5',
 mRNA sequence.
ACCESSION BI856603
VERSION BI856603.1 GI:15997350
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 775)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: DCTD/DTF
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLNL2005 row: f column: 10
 High quality sequence stop: 508.

FEATURES
 source

1..775
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5394441"
 /clone_lib="NIH_MGC_87"
 /tissue_type="mammary adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 1.383 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."
 BASE COUNT 197 a 201 c 217 g 160 t

Query Match
 Best Local Similarity 1.4%; Score 21; DB 10; Length 775;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 895 ggaatttagcgaggaatttc 915

```

|||||GGAAATTTAGCGGAGAAATTC 157
Db 177

RESULT 7
AI068982 801 bp mRNA linear EST 09-DEC-1999
LOCUS mgae0004dh12f Magnaporthe grisea Appressorium Stage cDNA Library
DEFINITION Magnaporthe grisea cDNA clone mgae0004dh12f 5', mRNA sequence.
ACCESSION AI068982
VERSION AI068982.1 GI:3391957
KEYWORDS EST.
SOURCE Magnaporthe grisea.
ORGANISM Magnaporthe grisea
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
REFERENCE 1 (bases 1 to 801)
AUTHORS Choi,W., Fang,E., Sasinowski,M., Wing,R. and Dean,R.A.
TITLE Expressed sequence characterization during appressorium formation
in rice blast fungus, Magnaporthe grisea
JOURNAL Unpublished (1998)
COMMENT Contact: Dean,R.A.
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson University, Clemson, SC 29634, USA
Tel: 864 656 5737
Fax: 864 656 4293
Email: rdean@clemson.edu
Seq primer: T3 primer (AATTAACTCTACTAAAGG)
High quality sequence stop: 328.
FEATURES
source
1. .801
/organism="Magnaporthe grisea"
/strain="70-15"
/db_xref="taxon:148305"
/clone_lib="mgae0004dh12f"
/clone="mgae0004dh12f"
/library="Library"
/dev_stage="Germinated conidia on appressorium-inductive
surface"
/notes="Vector: pBlueScriptII SK(+) Vector; Site_1: EcoRI;
Site_2: XhoI; The appressorium formation-specific cDNA
library was constructed from conidia germinated for 5-8
hr on an inductive surface. The library has an average
insert size of 1.5 kbp."
BASE COUNT 202 a 223 c 203 g 168 t
ORIGIN

Query Match 1.4%; Score 21; DB 9; Length 801;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1159 gagagagatctttggcccgctc 1179
|||||GGAGGATCTTTGGCCCGCTC 590
Db 570

RESULT 8
BI258465/c 852 bp mRNA linear EST 17-JUL-2001
LOCUS 602972456f1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5112037 5',
DEFINITION mRNA sequence.
ACCESSION BI258465
VERSION BI258465.1 GI:14814836
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 852)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11271 row: o column: 14
High quality sequence stop: 832.
FEATURES
source
1. .852
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5112037"
/clone_lib="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/notes="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library prepared by Life
Technologies."
BASE COUNT 199 a 219 c 254 g 180 t
ORIGIN

Query Match 1.4%; Score 21; DB 10; Length 852;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 895 ggaaatttagcgagaaattc 915
|||||GGAAATTTAGCGGAGAAATTC 139
Db 159

RESULT 9
BF035820/c 930 bp mRNA linear EST 20-OCT-2000
LOCUS 601458344f1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3861923 5',
DEFINITION mRNA sequence.
ACCESSION BF035820
VERSION BF035820.1 GI:10743560
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 930)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/Drp
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9599 row: g column: 12
High quality sequence stop: 9
High quality sequence stop: 607.
FEATURES
source
1. .930
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3861923"
/clone_lib="NIH_MGC_66"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"

```



```

/notes="Organ: ovary; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT      222 a   295 c   251 g   162 t
ORIGIN

Query Match      1.4%; Score 21; DB 10; Length 930;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 895 ggaaatttagcgagaaattc 915
|||||
Db 125 GGAAATTTAGCGGAGAAATTC 105

RESULT 10
AL536113
LOCUS
DEFINITION AL536113 Ltr1_FL013_Fbrn1 Homo sapiens cDNA clone CS0DF022YK12 5
prime, mRNA sequence.
ACCESSION AL536113
VERSION AL536113
KEYWORDS AL536113.1 GI:12799606
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 938)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..938
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF022YK12"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/lab_host="pH108"
/notes="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
filiang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT      0 a   526 c   291 g   120 t   1 others
ORIGIN

Query Match      1.4%; Score 21; DB 9; Length 938;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 505 gcccccgcctcgccgcggc 525
|||||
Db 263 GCCCCCGCGCTCGCGCGCGC 283

RESULT 11
BL457385/c
LOCUS
DEFINITION 603185515F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:5258306 5',

```

```

mRNA sequence.
BI457385
GI:15248041
EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1243)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1869 row: n column: 03
High quality sequence stop: 263.
Location/Qualifiers
1..1243
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5258306"
/clone_lib="NIH_MGC_42"
/tissue_type="epithelioid carcinoma cell line"
/lab_host="pH108 (phage-resistant)"
/notes="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(s). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT      379 a   355 c   318 g   191 t
ORIGIN

Query Match      1.4%; Score 21; DB 10; Length 1243;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 895 ggaaatttagcgagaaattc 915
|||||
Db 214 GGAAATTTAGCGGAGAAATTC 194

RESULT 12
BG816128/c
LOCUS
DEFINITION BG816128 y1 NICHD XGC Emb2 Xenopus laevis cDNA clone IMAGE:4406618
5', similar to SW:OXYB_HUMAN P22059 OXYSTEROL-BINDING PROTEIN. ;
mRNA sequence.
BASE COUNT      534 bp   mRNA   linear   EST 22-MAY-2001
ORGANISM BG816128
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 534)
AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterston,R. and Wilton,R.
WashU Xenopus EST project, 1999

```

JOURNAL
COMMENT

Unpublished (1999)
Other ESTs: dac07h02.x1
Contact: Sandy Clifton, ph.D.
Washu Xenopus EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by Life Technologies. DNA Sequencing by:
Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
High quality sequence stop: 434.

FEATURES
source

1. .534
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:406618"
/clone_lib="NICHG XGC Emb2"
/tissue_type="embryo, stage 17/19"
/lab_host="DH10B (phage-resistant)"
/note="Vector: pCMV-SPORT6; Site.1: NotI; Site.2: SalI;
Cloned unidirectionally. Primer: Oligo dt. Average insert
size 2.1 kb. Constructed by Life Technologies."
BASE COUNT 158 a 102 c 129 g 145 t
ORIGIN
Query Match 1.4%; Score 20; DB 10; Length 534;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 524 gctgctcttcctcatcttcaaaa 543
|||||
Db 190 GCTGCTCTTCATCTTCAAA 171

RESULT 13
AI293860/c

LOCUS
DEFINITION AI293860 572 bp mRNA linear EST 19-APR-2001
LP07108.3prime LP Drosophila melanogaster larval-early pupal pot2
Drosophila melanogaster cDNA clone LP07108 3prime, mRNA sequence.

ACCESSION
AI293860

VERSION AI293860.1 GI:3943267
KEYWORDS EST.
SOURCE fruit fly.

ORGANISM

Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BDGP/HIMI Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP

TITLE
JOURNAL

COMMENT Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Based upon the presence of a XhoI site followed by a run of 14 or
more T residues at the beginning of the sequence, this clone
probably contains an inverted insert. The sequence has been trimmed
and the T residues removed.
Plate: 71 row: A column: 8
High quality sequence stop: 467.
Location/Qualifiers
1. .572
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="LP07108"

FEATURES
source

1. .572
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="LP07108"

/clone_lib="LP Drosophila melanogaster larval-early pupal
pot2"
/sex="male and female"
/dev_stage="larvae-pupae"
/lab_host="DH5-alpha"
/note="Organ: whole body; Vector: pot2; Site.1: EcoRI;
Site.2: XhoI: sized fractionated cDNAs were directly
ligated into pot2. Plasmid cDNA library."

BASE COUNT 151 a 154 c 130 g 137 t
ORIGIN

Query Match 1.4%; Score 20; DB 9; Length 572;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1159 gaggagatctttggcccgct 1178
|||||
Db 402 GAGGAGATCTTTGGCCCGCT 383

RESULT 14
BJ060784/c

LOCUS
DEFINITION BJ060784 582 bp mRNA linear EST 11-DEC-2001
laevis cDNA clone XL066n04 5', mRNA sequence.

ACCESSION
BJ060784

VERSION BJ060784.1 GI:17500525
KEYWORDS EST.
SOURCE African clawed frog.

ORGANISM

Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 582)

REFERENCE

AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-I,T. and Kohara
Y.
Expressed genes in X. laevis embryo
Unpublished (2001)
Contact: Tadasu Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
source

Location/Qualifiers
1. .582
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="XL066n04"
/clone_lib="NIBB Mochii normalized Xenopus tailbud
library"
/tissue_type="whole embryo"
/dev_stage="stage 25"
BASE COUNT 171 a 115 c 124 g 170 t 2 others
ORIGIN

Query Match 1.4%; Score 20; DB 10; Length 582;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 524 gctgctcttcctcatcttcaaaa 543
|||||
Db 68 GCTGCTCTTCATCTTCAAA 49

RESULT 15
AI259976

LOCUS
DEFINITION AI259976 583 bp mRNA linear EST 19-APR-2001
LP03541.5prime LP Drosophila melanogaster larval-early pupal pot2

Search completed: April 27, 2002, 02:58:44
Job time: 13436 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 07:43:52 ; Search time 95.74 Seconds
(without alignments)
2809.045 Million cell updates/sec

Title: US-09-882-694A-4
Perfect score: 1464
Sequence: 1 atggtcttcgcctgaacga.....tccacgtcaacatagatga 1464.

Scoring table: OLIGO_NWC
Gapop 60.0 , Gapext 60.0

Searched: 139433 seqs, 91850399 residues

Word size : 0

Total number of hits satisfying chosen parameters: 278866

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending_Patents_NA_New.*
1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------------------|--------------------|
| C 1 | 18 | 1.2 | 33965 | US-10-105-299-12544 | Sequence 12544, A |
| C 2 | 17 | 1.2 | 259 | US-09-975-254-13294 | Sequence 13294, A |
| 3 | 17 | 1.2 | 569 | US-10-102-524-206 | Sequence 206, App |
| 4 | 17 | 1.2 | 961 | US-10-105-299-11655 | Sequence 11655, A |
| 5 | 17 | 1.2 | 964 | US-10-105-299-2100 | Sequence 2100, App |
| 6 | 17 | 1.2 | 3725 | US-10-116-802-334 | Sequence 334, App |
| C 7 | 17 | 1.2 | 32493 | US-09-539-698A-6 | Sequence 6, Appli |
| 8 | 17 | 1.2 | 39339 | US-10-105-299-7119 | Sequence 7119, Ap |
| 9 | 16 | 1.1 | 99 | US-10-106-698-3137 | Sequence 3137, Ap |
| 10 | 16 | 1.1 | 197 | US-10-097-105-1475 | Sequence 1475, Ap |
| 11 | 16 | 1.1 | 218 | US-09-540-2108-3171 | Sequence 3171, Ap |
| 12 | 16 | 1.1 | 235 | US-09-540-2108-16680 | Sequence 16680, A |
| 13 | 16 | 1.1 | 246 | US-09-975-254-13104 | Sequence 13104, A |
| 14 | 16 | 1.1 | 248 | US-09-975-254-18042 | Sequence 18042, A |
| C 15 | 16 | 1.1 | 251 | US-09-540-2108-205 | Sequence 205, App |
| C 16 | 16 | 1.1 | 412 | US-60-365-384-462 | Sequence 462, App |
| C 17 | 16 | 1.1 | 412 | US-60-365-384-463 | Sequence 463, App |
| 18 | 16 | 1.1 | 807 | US-10-107-096-1 | Sequence 1, Appli |
| 19 | 16 | 1.1 | 978 | US-09-540-2098-4283 | Sequence 4283, Ap |
| 20 | 16 | 1.1 | 1139 | US-10-102-806-122 | Sequence 122, App |
| C 21 | 16 | 1.1 | 1335 | US-09-312-2838-76 | Sequence 76, Appl |
| C 22 | 16 | 1.1 | 1335 | US-09-312-2838-261 | Sequence 261, App |
| 23 | 16 | 1.1 | 1516 | US-60-365-384-120 | Sequence 120, App |
| 24 | 16 | 1.1 | 1576 | US-60-365-384-121 | Sequence 121, App |
| 25 | 16 | 1.1 | 1609 | US-09-896-908-1 | Sequence 1, Appli |
| 26 | 16 | 1.1 | 1647 | US-09-896-908-14 | Sequence 14, Appl |

| | | | | | | |
|------|----|-----|-------|---|---------------------|-------------------|
| 27 | 16 | 1.1 | 2061 | 5 | US-09-540-209B-817 | Sequence 817, App |
| 28 | 16 | 1.1 | 2898 | 6 | US-10-029-397A-52 | Sequence 52, Appl |
| 29 | 16 | 1.1 | 2900 | 6 | US-10-029-397A-55 | Sequence 55, Appl |
| 30 | 16 | 1.1 | 2901 | 6 | US-10-029-397A-57 | Sequence 57, Appl |
| 31 | 16 | 1.1 | 2902 | 6 | US-10-029-397A-56 | Sequence 56, Appl |
| 32 | 16 | 1.1 | 2912 | 6 | US-10-029-397A-51 | Sequence 51, Appl |
| 33 | 16 | 1.1 | 2922 | 6 | US-10-029-397A-50 | Sequence 50, Appl |
| 34 | 16 | 1.1 | 2923 | 6 | US-10-029-397A-54 | Sequence 54, Appl |
| 35 | 16 | 1.1 | 2932 | 6 | US-10-029-397A-53 | Sequence 53, Appl |
| 36 | 16 | 1.1 | 5196 | 5 | US-09-630-630B-28 | Sequence 28, Appl |
| C 37 | 16 | 1.1 | 7025 | 6 | US-10-105-299-9628 | Sequence 9628, Ap |
| C 38 | 16 | 1.1 | 9525 | 6 | US-10-105-299-8161 | Sequence 8161, Ap |
| C 39 | 16 | 1.1 | 9525 | 6 | US-10-105-299-8162 | Sequence 8162, Ap |
| C 40 | 16 | 1.1 | 9525 | 6 | US-10-105-299-8172 | Sequence 8172, Ap |
| C 41 | 16 | 1.1 | 9525 | 6 | US-10-105-299-8173 | Sequence 8173, Ap |
| C 42 | 16 | 1.1 | 11538 | 6 | US-10-105-299-12735 | Sequence 12735, A |
| C 43 | 16 | 1.1 | 12390 | 6 | US-10-105-299-8164 | Sequence 8164, Ap |
| C 44 | 16 | 1.1 | 12390 | 6 | US-10-105-299-8175 | Sequence 8175, Ap |
| 45 | 16 | 1.1 | 13003 | 6 | US-10-112-699-2911 | Sequence 2911, Ap |

ALIGNMENTS

RESULT 1
US-10-105-299-12544/c
; Sequence 12544, Application US/10105299
; GENERAL INFORMATION:
; APPLICANT: Rosen, et. al
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS950
; CURRENT APPLICATION NUMBER: US/10/105,299
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ ID NOS: 15197
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12544
; LENGTH: 33965
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-105-299-12544

Query Match 1.2%; Score 18; DB 6; Length 33965;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 529 tcttcattcttcaagcc 546
|||||
Db 13262 TCCTTCATCTTCAAGCC 13245

RESULT 2
US-09-975-254-13294/c
; Sequence 13294, Application US/09975254
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Heck, Gregory R.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(15309)B
; CURRENT APPLICATION NUMBER: US/09/975,254
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US/09/263,191
; PRIOR FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 31255
; SEQ ID NO 13294
; LENGTH: 259
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 700907504H1
US-09-975-254-13294

```

Query Match      1.2%; Score 17; DB 5; Length 259;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1135 aaaccaggcagcaaat 1151
Db 75 AAACAGGCGACGCAAAAT 59

RESULT 3
US-10-102-524-206
; Sequence 206, Application US/10102524
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Gaiger, Alexander
; APPLICANT: Gordon, Brian
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF KIDNEY CANCER
; FILE REFERENCE: 210121.572
; CURRENT APPLICATION NUMBER: US/10/102,524
; CURRENT FILING DATE: 2002-03-19
; NUMBER OF SEQ ID NOS: 1863
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 206
; LENGTH: 569
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-102-524-206

Query Match      1.2%; Score 17; DB 6; Length 569;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 641 gagtgcgggtgaagca 657
Db 229 gagtgcgggtgaagca 245

RESULT 4
US-10-105-299-11655
; Sequence 11655, Application US/10105299
; GENERAL INFORMATION:
; APPLICANT: Rosen, et. al
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS950
; CURRENT APPLICATION NUMBER: US/10/105,299
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ ID NOS: 15197
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11655
; LENGTH: 961
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-105-299-11655

Query Match      1.2%; Score 17; DB 6; Length 961;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 118 actgatgttcacgtgac 134
Db 323 actgatgttcacgtgac 339

RESULT 5
US-10-105-299-2100
; Sequence 2100, Application US/10105299
; GENERAL INFORMATION:
; APPLICANT: Rosen, et. al
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS950
; CURRENT APPLICATION NUMBER: US/10/105,299
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ ID NOS: 15197
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2100
; LENGTH: 964
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (883)..(883)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-105-299-2100

Query Match      1.2%; Score 17; DB 6; Length 964;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 118 actgatgttcacgtgac 134
Db 321 actgatgttcacgtgac 337

RESULT 6
US-10-116-802-334
; Sequence 334, Application US/10116802
; GENERAL INFORMATION:
; APPLICANT: Amy Lasek
; TITLE OF INVENTION: GENES EXPRESSED IN LUNG CANCER
; FILE REFERENCE: PA-0045 US
; CURRENT APPLICATION NUMBER: US/10/116,802
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: 60/281,593
; PRIOR FILING DATE: 2001-04-04
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: PERL Program
; SEQ ID NO 334
; LENGTH: 3725
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 474926.11
US-10-116-802-334

Query Match      1.2%; Score 17; DB 6; Length 3725;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 530 ccttcattcttcaagcc 546
Db 167 ccttcattcttcaagcc 183

RESULT 7
US-09-539-698A-6/c
; Sequence 6, Application US/09539698A
; GENERAL INFORMATION:
; APPLICANT: GenStar Therapeutics Corporation
; APPLICANT: Alemany, Ramon
; APPLICANT: Fang, Xiangming
; APPLICANT: Zhang, Wei-Wei
; APPLICANT: Robert, Sobol
; TITLE OF INVENTION: Complementary-Adenoviral Vector System
; FILE REFERENCE: 97-087-B
US-09-539-698A-6/c

```

```
; CURRENT APPLICATION NUMBER: US/09/539,698A
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 08/797,160
; PRIOR FILING DATE: 1997-02-10
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(32493)
; OTHER INFORMATION: GIV8053 plasmid vector
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (10461)..(16291)
; OTHER INFORMATION: complement (10461..16291), PSA promoter/enhancer sequence
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: (7276)..(7423)
; OTHER INFORMATION: polyA signal sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (7441)..(10440)
; OTHER INFORMATION: Elae1b sequence coding sequence in 10440..7441 orientation
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (16742)..(17138)
; OTHER INFORMATION: RSV LTR
; FEATURE:
; NAME/KEY: Intron
; LOCATION: (17164)..(17284)
; OTHER INFORMATION: SV40 intron
; FEATURE:
; NAME/KEY: gene
; LOCATION: (17395)..(18262)
; OTHER INFORMATION: human B7-1 cDNA sequence
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: (18292)..(18490)
; OTHER INFORMATION: SV40 poly A signal sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (18523)..(19785)
; OTHER INFORMATION: EF promoter sequence
; FEATURE:
; NAME/KEY: gene
; LOCATION: (19805)..(20305)
; OTHER INFORMATION: human interferon gamma (hIFNg) cDNA sequence
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: (20712)..(21113)
; OTHER INFORMATION: polyA signal sequence
; FEATURE:
; NAME/KEY: gene
; LOCATION: (23662)..(24559)
; OTHER INFORMATION: EGFP sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (24567)..(32050)
; OTHER INFORMATION: AFP D33 sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (487)..(7251)
; OTHER INFORMATION: Alb uB5 sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (22957)..(23646)
; OTHER INFORMATION: SV40 promoter and first intron
; FEATURE:
; NAME/KEY: LTR
; LOCATION: (1)..(100)
```

```
; OTHER INFORMATION: adenovirus (Ad) ITR5'
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (236)..(373)
; OTHER INFORMATION: packaging signal
; FEATURE:
; NAME/KEY: LTR
; LOCATION: (32393)..(32493)
; OTHER INFORMATION: adenovirus (Ad) ITR3'
; US-09-539-698A-6

Query Match 1.2%; Score 17; DB 5; Length 32493;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 188 agggcccatggaagaag 204
      |||||
DB 20668 AGGGCCCATGGAAGAAG 20652

RESULT 8
US-10-105-299-7119
; Sequence 7119, Application US/10105299
; GENERAL INFORMATION:
; APPLICANT: Rosen, et. al
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS950
; CURRENT APPLICATION NUMBER: US/10/105,299
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ ID NOS: 15197
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7119
; LENGTH: 39339
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-105-299-7119

Query Match 1.2%; Score 17; DB 6; Length 39339;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 599 ccggattccctctgga 615
      |||||
DB 36 ccggattccctctgga 52

RESULT 9
US-10-106-698-3137
; Sequence 3137, Application US/10106698
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypep
; FILE REFERENCE: PA00521
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 3137
; LENGTH: 99
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (7)..(7)
```

```

; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (27)..(27)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (48)..(48)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (50)..(50)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (53)..(53)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (63)..(63)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (70)..(70)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (90)..(90)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (92)..(92)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-3137

```

Query Match 1.1%; Score 16; DB 6; Length 99;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 16; Conservative 0; Mismatches 0; Indels

Qy 839 atttctcaaaattcgg 854
 | | | | | | | | | |
Db 31 atttctcaaaattcgg 46

```

RESULT 10
US-10-097-105-1475
; SEQUENCE 1475, Application US/10097105
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: King, Gordon E.
; APPLICANT: Secrist, Heather
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF COLON CANCER
; FILE REFERENCE: 210121.504C1
; CURRENT APPLICATION NUMBER: US/10/097.105
; CURRENT FILING DATE: 2002-03-13
; NUMBER OF SEQ ID NOS: 1562
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1475
; LENGTH: 197
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-097-105-1475

```

```
Query Match      1.1%; Score 16; DB 6; Length 197;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 16; Conservative 0; Mismatches 0; Indels
```

Qy 733 acactcaagtccaaca 748
 Dd 123 acactcaagtccaaca 138

RESULT 11
US-09-540-210B-3171
; Sequence 3171, Application US/09540210B
; GENERAL INFORMATION:

? PRIOR FILING DATE: February 13, 1997
 ? PRIOR APPLICATION NUMBER: 09/035,172
 ? PRIOR FILING DATE: March 4, 1998
 ? PRIOR APPLICATION NUMBER: 60/040,431
 ? PRIOR FILING DATE: March 5, 1997
 ? PRIOR APPLICATION NUMBER: 09/041,894
 ? PRIOR FILING DATE: March 12, 1998
 ? PRIOR APPLICATION NUMBER: 60/040,199
 ? PRIOR FILING DATE: March 14, 1997
 ? PRIOR APPLICATION NUMBER: 09/050,817
 ? PRIOR FILING DATE: March 30, 1998
 ? PRIOR APPLICATION NUMBER: 60/043,792
 ? PRIOR FILING DATE: April 11, 1997
 ? PRIOR APPLICATION NUMBER: 09/074,999
 ? PRIOR FILING DATE: May 8, 1998
 ? PRIOR APPLICATION NUMBER: 60/048,431
 ? PRIOR FILING DATE: May 29, 1997
 ? PRIOR APPLICATION NUMBER: 09/107,592
 ? PRIOR FILING DATE: June 30, 1998
 ? PRIOR APPLICATION NUMBER: 60/052,751
 ? PRIOR FILING DATE: July 1, 1997
 ? PRIOR APPLICATION NUMBER: 09/094,079
 ? PRIOR FILING DATE: June 9, 1998
 ? PRIOR APPLICATION NUMBER: 60/049,975
 ? PRIOR FILING DATE: June 13, 1997
 ? NUMBER OF SEQ ID NOS: 35654
 ? SOFTWARE: PERL Program
 ? SEQ ID NO 3171
 ? LENGTH: 218
 ? TYPE: DNA
 ? ORGANISM: Homo sapiens
 ? FEATURE:
 ? NAME/KEY: misc.feature
 ? OTHER INFORMATION: Incyte ID No: hu00348038
 US-09-540-210B-3171

Query Match 1.18; Score 16; DB 5; Length 218;

Best Local Similarity 100.08; Pred. No. 33; Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 733 acactcaagtccaaca 748
 |||||
 Db 72 acactcaagtccaaca 87

RESULT 12
 US-09-540-210B-16680
 ? Sequence 16680, Application US/09540210B
 ? GENERAL INFORMATION:
 ? APPLICANT: Sellhammer, Jeffrey J.
 ? APPLICANT: Deleageane, Angelo M.
 ? APPLICANT: Stuart, Susan G.
 ? APPLICANT: Stuve, Laura L.
 ? APPLICANT: Mullahy, Sara J.
 ? APPLICANT: Naughton, Rebecca E.
 ? TITLE OF INVENTION: POLYNUCLEOTIDES OF URINARY TRACT TISSUE
 ? FILE REFERENCE: PD-1037 CIP
 ? CURRENT APPLICATION NUMBER: US/09/540,210B
 ? PRIOR FILING DATE: 2002-04-03
 ? PRIOR APPLICATION NUMBER: 08/972,899
 ? PRIOR FILING DATE: November 18, 1997
 ? PRIOR APPLICATION NUMBER: 08/395,244
 ? PRIOR FILING DATE: February 27, 1995
 ? PRIOR APPLICATION NUMBER: 08/722,922
 ? PRIOR FILING DATE: September 27, 1996
 ? PRIOR APPLICATION NUMBER: 60/005,526
 ? PRIOR FILING DATE: September 29, 1995
 ? PRIOR APPLICATION NUMBER: 08/824,029
 ? PRIOR FILING DATE: March 25, 1997
 ? PRIOR APPLICATION NUMBER: 60/014,010
 ? PRIOR FILING DATE: March 25, 1996
 ? PRIOR APPLICATION NUMBER: 08/826,847

? PRIOR FILING DATE: April 10, 1997
 ? PRIOR APPLICATION NUMBER: 60/015,533
 ? PRIOR FILING DATE: April 10, 1996
 ? PRIOR APPLICATION NUMBER: 08/903,555
 ? PRIOR FILING DATE: July 31, 1997
 ? PRIOR APPLICATION NUMBER: 60/023,308
 ? PRIOR FILING DATE: July 31, 1996
 ? PRIOR APPLICATION NUMBER: 08/862,178
 ? PRIOR FILING DATE: May 22, 1997
 ? PRIOR APPLICATION NUMBER: 60/018,217
 ? PRIOR FILING DATE: May 23, 1996
 ? PRIOR APPLICATION NUMBER: 08/881,589
 ? PRIOR FILING DATE: June 24, 1997
 ? PRIOR APPLICATION NUMBER: 60/021,275
 ? PRIOR FILING DATE: June 25, 1996
 ? PRIOR APPLICATION NUMBER: 08/903,802
 ? PRIOR FILING DATE: July 31, 1997
 ? PRIOR APPLICATION NUMBER: 60/023,308
 ? PRIOR FILING DATE: July 31, 1996
 ? PRIOR APPLICATION NUMBER: 08/905,881
 ? PRIOR FILING DATE: August 1, 1997
 ? PRIOR APPLICATION NUMBER: 60/025,204
 ? PRIOR FILING DATE: August 1, 1996
 ? PRIOR APPLICATION NUMBER: 08/903,471
 ? PRIOR FILING DATE: July 30, 1997
 ? PRIOR APPLICATION NUMBER: 60/025,478
 ? PRIOR FILING DATE: July 31, 1996
 ? PRIOR APPLICATION NUMBER: 08/903,556
 ? PRIOR FILING DATE: July 31, 1997
 ? PRIOR APPLICATION NUMBER: 60/025,217
 ? PRIOR FILING DATE: August 22, 1996
 ? PRIOR APPLICATION NUMBER: 08/937,142
 ? PRIOR FILING DATE: September 23, 1997
 ? PRIOR APPLICATION NUMBER: 60/026,598
 ? PRIOR FILING DATE: September 24, 1996
 ? PRIOR APPLICATION NUMBER: 08/960,746
 ? PRIOR FILING DATE: October 29, 1997
 ? PRIOR APPLICATION NUMBER: 60/030,144
 ? PRIOR FILING DATE: October 30, 1996
 ? PRIOR APPLICATION NUMBER: 08/826,847
 ? PRIOR FILING DATE: April 10, 1997
 ? PRIOR APPLICATION NUMBER: 60/015,533
 ? PRIOR FILING DATE: April 10, 1996
 ? PRIOR APPLICATION NUMBER: 08/755,524
 ? PRIOR FILING DATE: November 22, 1996
 ? PRIOR APPLICATION NUMBER: 60/007,495
 ? PRIOR FILING DATE: November 22, 1995
 ? PRIOR APPLICATION NUMBER: 09/021,031
 ? PRIOR FILING DATE: February 10, 1998
 ? PRIOR APPLICATION NUMBER: 60/039,325
 ? PRIOR FILING DATE: February 13, 1997
 ? PRIOR APPLICATION NUMBER: 09/035,172
 ? PRIOR FILING DATE: March 4, 1998
 ? PRIOR APPLICATION NUMBER: 60/040,431
 ? PRIOR FILING DATE: March 5, 1997
 ? PRIOR APPLICATION NUMBER: 09/041,894
 ? PRIOR FILING DATE: March 12, 1998
 ? PRIOR APPLICATION NUMBER: 60/040,199
 ? PRIOR FILING DATE: March 14, 1997
 ? PRIOR APPLICATION NUMBER: 09/050,817
 ? PRIOR FILING DATE: March 30, 1998
 ? PRIOR APPLICATION NUMBER: 60/043,792
 ? PRIOR FILING DATE: April 11, 1997
 ? PRIOR APPLICATION NUMBER: 09/074,999
 ? PRIOR FILING DATE: May 8, 1998
 ? PRIOR APPLICATION NUMBER: 60/048,431
 ? PRIOR FILING DATE: May 29, 1997
 ? PRIOR APPLICATION NUMBER: 09/107,592
 ? PRIOR FILING DATE: June 30, 1998
 ? PRIOR APPLICATION NUMBER: 60/052,751
 ? PRIOR FILING DATE: July 1, 1997
 ? PRIOR APPLICATION NUMBER: 09/094,079

; PRIOR APPLICATION NUMBER: 60/049,975
; PRIOR FILING DATE: June 13, 1997
; NUMBER OF SEQ ID NOS: 35654
; SOFTWARE: PERL Program
; SEQ ID NO 16680
; LENGTH: 235
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID NO: hu00457807
; NAME/KEY: unsure
; LOCATION: 108, 115
; OTHER INFORMATION: a, t, c, g, or other
US-09-540-210B-16680

Query Match 1.1%; Score 16; DB 5; Length 235;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 733 acactcaagtcaca 748
|||||

Db 23 acactcaagtcaca 38

RESULT 13
US-09-975-254-13104
; Sequence 13104, Application US/09975254
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Heck, Gregory R.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15309)B
; CURRENT APPLICATION NUMBER: US/09/975,254
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US/09/263,191
; PRIOR FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 31255
; SEQ ID NO 13104
; LENGTH: 246
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 700907183H1
US-09-975-254-13104

Query Match 1.1%; Score 16; DB 5; Length 246;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 908 agaaattccatggagt 923
|||||

Db 169 agaaattccatggagt 184

RESULT 14
US-09-975-254-18042
; Sequence 18042, Application US/09975254
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Heck, Gregory R.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15309)B
; CURRENT APPLICATION NUMBER: US/09/975,254
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US/09/263,191
; PRIOR FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 31255

; SEQ ID NO 18042
; LENGTH: 248
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 700955062H1
US-09-975-254-18042

Query Match 1.1%; Score 16; DB 5; Length 248;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 908 agaaattccatggagt 923
|||||

Db 184 agaaattccatggagt 199

RESULT 15
US-09-540-210B-205/c
; Sequence 205, Application US/09540210B
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF URINARY TRACT TISSUE
; FILE REFERENCE: PD-1037 CIP
; CURRENT APPLICATION NUMBER: US/09/540,210B
; CURRENT FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 08/972,899
; PRIOR FILING DATE: November 18, 1997
; PRIOR APPLICATION NUMBER: 08/395,244
; PRIOR FILING DATE: February 27, 1995
; PRIOR APPLICATION NUMBER: 08/722,922
; PRIOR FILING DATE: September 27, 1996
; PRIOR APPLICATION NUMBER: 60/005,526
; PRIOR FILING DATE: September 29, 1995
; PRIOR APPLICATION NUMBER: 08/824,029
; PRIOR FILING DATE: March 25, 1997
; PRIOR APPLICATION NUMBER: 60/014,010
; PRIOR FILING DATE: March 25, 1996
; PRIOR APPLICATION NUMBER: 08/826,847
; PRIOR FILING DATE: April 10, 1997
; PRIOR APPLICATION NUMBER: 60/015,533
; PRIOR FILING DATE: April 10, 1996
; PRIOR APPLICATION NUMBER: 08/903,555
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/023,308
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/862,178
; PRIOR FILING DATE: May 22, 1997
; PRIOR APPLICATION NUMBER: 60/018,217
; PRIOR FILING DATE: May 23, 1996
; PRIOR APPLICATION NUMBER: 08/881,589
; PRIOR FILING DATE: June 24, 1997
; PRIOR APPLICATION NUMBER: 60/021,275
; PRIOR FILING DATE: June 25, 1996
; PRIOR APPLICATION NUMBER: 08/903,802
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/023,308
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/905,881
; PRIOR FILING DATE: August 1, 1997
; PRIOR APPLICATION NUMBER: 60/025,204
; PRIOR FILING DATE: August 1, 1996
; PRIOR APPLICATION NUMBER: 08/903,471
; PRIOR FILING DATE: July 30, 1997
; PRIOR APPLICATION NUMBER: 60/025,478
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/903,556
; PRIOR FILING DATE: July 31, 1997

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 07:41:43 ; Search time 8172.53 Seconds
(without alignments)
3875.604 Million cell updates/sec

Title: US-09-882-694A-4
Perfect score: 1464
Sequence: 1 aggtcttcgcctgacga.....tccacgtcaacatagagtga 1464

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 21979536 seqs, 10817449327 residues

Word size : 0

Total number of hits satisfying chosen parameters: 43959072

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending_Patents_NA_Main:*

1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*

2: /cgn2_6/ptodata/2/pna/US08_COMB.seq:*

3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*

4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*

5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*

6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*

7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*

8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*

9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*

10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*

11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*

12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*

13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*

14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*

15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*

16: /cgn2_6/ptodata/2/pna/US092_COMB.seq:*

17: /cgn2_6/ptodata/2/pna/US093_COMB.seq:*

18: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*

19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*

20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*

21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*

22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*

23: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*

24: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*

25: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*

26: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*

27: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*

28: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*

29: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*

30: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*

31: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*

32: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*

33: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*

34: /cgn2_6/ptodata/2/pna/US097D_COMB.seq:*

35: /cgn2_6/ptodata/2/pna/US097E_COMB.seq:*

36: /cgn2_6/ptodata/2/pna/US097F_COMB.seq:*

37: /cgn2_6/ptodata/2/pna/US097G_COMB.seq:*

38: /cgn2_6/ptodata/2/pna/US097H_COMB.seq:*

39: /cgn2_6/ptodata/2/pna/US097I_COMB.seq:*

40: /cgn2_6/ptodata/2/pna/US097J_COMB.seq:*

41: /cgn2_6/ptodata/2/pna/US097K_COMB.seq:*

42: /cgn2_6/ptodata/2/pna/US097L_COMB.seq:*

43: /cgn2_6/ptodata/2/pna/US097M_COMB.seq:*

44: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*

45: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*

46: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*

47: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*

48: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*

49: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*

50: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*

51: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*

52: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*

53: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*

54: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*

55: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*

56: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*

57: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*

58: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*

59: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*

60: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*

61: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*

62: /cgn2_6/ptodata/2/pna/US6023_COMB.seq:*

63: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*

64: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*

65: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*

66: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*

67: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*

68: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*

69: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*

70: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*

71: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*

72: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*

73: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*

74: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*

75: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 1464 | 100.0 | 1464 | 17 | US-09-351-224-4 |
| 2 | 1464 | 100.0 | 1464 | 17 | US-09-351-224E-4 |
| 3 | 1464 | 100.0 | 1464 | 17 | US-09-351-823-4 |
| 4 | 1464 | 100.0 | 1464 | 26 | US-09-677-488-4 |
| 5 | 1464 | 100.0 | 1464 | 26 | US-09-677-488A-4 |
| 6 | 1464 | 100.0 | 1464 | 26 | US-09-677-682-4 |
| 7 | 1464 | 100.0 | 1464 | 26 | US-09-677-682A-4 |
| 8 | 1464 | 100.0 | 1464 | 26 | US-09-677-682B-4 |
| 9 | 1464 | 100.0 | 1464 | 33 | US-09-882-694-4 |
| 10 | 1464 | 100.0 | 1464 | 33 | US-09-882-694A-4 |
| 11 | 22 | 1.5 | 12332 | 24 | US-09-620-392-55184 |
| 12 | 22 | 1.5 | 20140 | 28 | US-09-702-134-4827 |
| 13 | 22 | 1.5 | 20140 | 31 | US-09-815-264-82296 |
| 14 | 21 | 1.4 | 340 | 31 | US-09-804-730-13078 |
| 15 | 21 | 1.4 | 340 | 57 | US-60-189-657-13072 |
| 16 | 21 | 1.4 | 363 | 26 | US-09-675-7844-1380 |
| 17 | 21 | 1.4 | 471 | 23 | US-09-605-702-16631 |
| 18 | 21 | 1.4 | 538 | 23 | US-09-605-702-16630 |
| 19 | 21 | 1.4 | 648 | 19 | US-09-526-263A-51 |
| 20 | 21 | 1.4 | 657 | 26 | US-09-675-784A-6649 |
| 21 | 21 | 1.4 | 704 | 16 | US-09-270-767-11385 |
| 22 | 21 | 1.4 | 1068 | 18 | US-09-417-507-12226 |
| 23 | 21 | 1.4 | 1134 | 24 | US-09-417-507-12222 |
| 24 | 21 | 1.4 | 3103 | 24 | US-09-620-392-66768 |
| 25 | 21 | 1.4 | 28326 | 24 | US-09-620-392-67807 |
| 26 | 21 | 1.4 | 28326 | 28 | US-09-702-134-9191 |
| 27 | 21 | 1.4 | 28326 | 31 | US-09-815-264-63345 |
| 28 | 21 | 1.4 | 97086 | 64 | US-60-258-279-8 |
| 29 | 20 | 1.4 | 361 | 64 | US-60-253-654-4231 |
| 30 | 20 | 1.4 | 361 | 64 | US-60-255-592-4231 |
| 31 | 20 | 1.4 | 378 | 26 | US-09-667-188A-6695 |

; OTHER INFORMATION: aldehyde dehydrogenase , coding sequence
US-09-351-823-4

```
Query Match      100.0%; Score 1464; DB 17; Length 1464;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggtttcttcgctgacgaatacaagagtgaaactttcatcaacaataatgctctcc 60
    |||||
Db 1 atggtttcttcgctgacgaatacaagagtgaaactttcatcaacaataatgctctcc 60

QY 61 tcaaggggttcgagagataaactctcaagaacccgtggagcaatccaccgttgcacat 120
    |||||
Db 61 tcaaggggttcgagagataaactctcaagaacccgtggagcaatccaccgttgcacat 120

QY 121 gatgttcacgtggccaaacgcccagatgctgacagtgcaagtgacgcttcggtcaggcg 180
    |||||
Db 121 gatgttcacgtggccaaacgcccagatgctgacagtgcaagtgacgcttcggtcaggcg 180

QY 181 gtcaaaaaaggcccatggagaagttcacaggtgcacaacgcgcggcgctgcatgctaa 240
    |||||
Db 181 gtcaaaaaaggcccatggagaagttcacaggtgcacaacgcgcggcgctgcatgctaa 240

QY 241 ttccgagacctgcgagagaacgcgcgagaaagctcgtctgagtcgctgcccacc 300
    |||||
Db 241 ttccgagacctgcgagagaacgcgcgagaaagctcgtctgagtcgctgcccacc 300

QY 301 ggtagaccggtgctgatgatactaatcttcacattccaaaatggtctccgtgttcgc 360
    |||||
Db 301 ggtagaccggtgctgatgatactaatcttcacattccaaaatggtctccgtgttcgc 360

QY 361 tactatgcagggtggccgacagaatgcgcggaaagacctttccgaggaacacgcgaag 420
    |||||
Db 361 tactatgcagggtggccgacagaatgcgcggaaagacctttccgaggaacacgcgaag 420

QY 421 ccgaattgcgttacagccgatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
    |||||
Db 421 ccgaattgcgttacagccgatgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480

QY 481 ttctttacgtcgtggtggaatagaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
    |||||
Db 481 ttctttacgtcgtggtggaatagaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 540

QY 541 aaagctcggagaaaatccccgcgtggggtcttgggcctcgtcctctctctctcagaagcc 600
    |||||
Db 541 aaagctcggagaaaatccccgcgtggggtcttgggcctcgtcctctctctctcagaagcc 600

QY 601 ggattccctcctggagtcgtcagttcctcactggagcacgagtgacgggtgaagcattg 660
    |||||
Db 601 ggattccctcctggagtcgtcagttcctcactggagcacgagtgacgggtgaagcattg 660

QY 661 ggcgtcacatggacattcgaagatcagcttcacaaagatctgtcggcggtggcgcgcc 720
    |||||
Db 661 ggcgtcacatggacattcgaagatcagcttcacaaagatctgtcggcggtggcgcgcc 720

QY 721 gtcaagcaagcaacactcaagtccaacatgaagcgcgtcactctagaactgggggaaag 780
    |||||
Db 721 gtcaagcaagcaacactcaagtccaacatgaagcgcgtcactctagaactgggggaaag 780

QY 781 ccaacctcgtttcaacaaagctcctcgcgaacgagtcggggaatcgaagaagat 840
    |||||
Db 781 ccaacctcgtttcaacaaagctcctcgcgaacgagtcggggaatcgaagaagat 840

QY 841 ttctcaaaattcggggcaaatgggttccccccctcctgtttgtcgtagtcgaatggggaat 900
    |||||
Db 841 ttctcaaaattcggggcaaatgggttccccccctcctgtttgtcgtagtcgaatggggaat 900

QY 901 ttacgggaataattccatgagtcgctcagctcatttgaagctgtcaagatggtct 960
    |||||
Db 901 ttacgggaataattccatgagtcgctcagctcatttgaagctgtcaagatggtct 960

QY 961 ggcgaagaacccattggaacccaagagacgcatgggtccctcgtcgaagaagtcaccagtac 1020
    |||||
```

```
Db 961 ggcgaagaacccattggaacccaagagagcgcacgttcccttcgtcgaagaagtcaccagtac 1020
    |||||
QY 1021 gacagagctcttgggttaacattgacgcttggcaagagataccgcgcgcgcgcgcgcgcgc 1080
    |||||
Db 1021 gacagagctcttgggttaacattgacgcttggcaagagataccgcgcgcgcgcgcgcgcgc 1080
    |||||
QY 1081 ggtagaagaaggcgcaagaaggattcgcgattgaacgcgacgatatattgtcaatcccaaaaca 1140
    |||||
Db 1081 ggtagaagaaggcgcaagaaggattcgcgattgaacgcgacgatatattgtcaatcccaaaaca 1140
    |||||
QY 1141 ggcagcaaaatttgggttgagagagatctttggcccccgtcttgcattaaagacgttcaag 1200
    |||||
Db 1141 ggcagcaaaatttgggttgagagagatctttggcccccgtcttgcattaaagacgttcaag 1200
    |||||
QY 1201 acggaagaagaggccattgagattgccaatgacacgacttatgggtagcctcgtcatt 1260
    |||||
Db 1201 acggaagaagaggccattgagattgccaatgacacgacttatgggtagcctcgtcatt 1260
    |||||
QY 1261 tataccaaatcttcaaacagagggtctcgtgtctcgttcggtcgcgcgcgcgcgcgcgcgcgc 1320
    |||||
Db 1261 tataccaaatcttcaaacagagggtctcgtgtctcgttcggtcgcgcgcgcgcgcgcgcgcgc 1320
    |||||
QY 1321 tcgatacaacttccccctttatccccgcgagacacaaactcctgttggcggcgcgcgcgcgcgcgc 1380
    |||||
Db 1321 tcgatacaacttccccctttatccccgcgagacacaaactcctgttggcggcgcgcgcgcgcgcgc 1380
    |||||
QY 1381 ggcctagcgacagagactagggcgaagaaggctcgaaggcgtacttggagcccaagaccatt 1440
    |||||
Db 1381 ggcctagcgacagagactagggcgaagaaggctcgaaggcgtacttggagcccaagaccatt 1440
    |||||
QY 1441 aatatccacgccaacatagagtga 1464
    |||||
Db 1441 aatatccacgccaacatagagtga 1464
    |||||
```

RESULT 4

```
US-09-677-488-4
; Sequence 4, Application US/09677488
; GENERAL INFORMATION:
; APPLICANT: Duwick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fummonisin Detoxification
; FILE REFERENCE: 35718/204100 (5718-111A)
; CURRENT APPLICATION NUMBER: US/09/677,488
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: aldehyde dehydrogenase , fully spliced cDNA
US-09-677-488-4
```

```
Query Match      100.0%; Score 1464; DB 26; Length 1464;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 atggtttcttcgctgacgaatacaagagtgaaactttcatcaacaataatgctctcc 60
    |||||
Db 1 atggtttcttcgctgacgaatacaagagtgaaactttcatcaacaataatgctctcc 60
    |||||
QY 61 tccaaggggttcgagagattaaactcgaacccgttgggacgaatccaccgttgcacat 120
    |||||
Db 61 tccaaggggttcgagagattaaactcgaacccgttgggacgaatccaccgttgcacat 120
    |||||
QY 121 gatgttcacgtggccaaacgcccagatgctgacagtcagtcagcgtctccgttcaggcg 180
    |||||
Db 121 gatgttcacgtggccaaacgcccagatgctgacagtcagtcagcgtctccgttcaggcg 180
    |||||
```



```
QY 181 gtcaaaagggccatgggaagttcacaggtgcacaacgcgagcggtgcatgcttaag 240
Db 181 gtcaaaagggccatgggaagttcacaggtgcacaacgcgagcggtgcatgcttaag 240
QY 241 ttgcggagactgcgcgagaaaccccgagagctgcgtctgagagctgctcccaac 300
Db 241 ttgcggagactgcgcgagaaaccccgagagctgcgtctgagagctgctcccaac 300
QY 301 ggtagaccggttcgatgatactactatttcacattcccaaatggtctcgttttgcg 360
Db 301 ggtagaccggttcgatgatactactatttcacattcccaaatggtctcgttttgcg 360
QY 361 tactatgcaggttcggccgacagatcccggaagactctcccgaggaacacgcgaag 420
Db 361 tactatgcaggttcggccgacagatcccggaagactctcccgaggaacacgcgaag 420
QY 421 cegaattggcgttaacagccgattggggtgtgtctggtattgcaagctggaacgact 480
Db 421 cegaattggcgttaacagccgattggggtgtgtctggtattgcaagctggaacgact 480
QY 481 ttctttacgttcggttcggaagatagcccccgcctcgcgcgcgtctctctcatcttc 540
Db 481 ttctttacgttcggttcggaagatagcccccgcctcgcgcgcgtctctctcatcttc 540
QY 541 aaagctcggagaaatcccccgtggggttctgggctcgtctctcttcgcagaagcc 600
Db 541 aaagctcggagaaatcccccgtggggttctgggctcgtctctcttcgcagaagcc 600
QY 601 ggaattcccttcgagctgcagttctcactcagacagacagtcgagggcgagacattg 660
Db 601 ggaattcccttcgagctgcagttctcactcagacagacagtcgagggcgagacattg 660
QY 661 gctcgcacatggacattgcgaagatcagcttcacaagatctgtcggcggtggcgcgcc 720
Db 661 gctcgcacatggacattgcgaagatcagcttcacaagatctgtcggcggtggcgcgcc 720
QY 721 gtcaagcaagcaactcaagtcacaatgaagcgcgtcactctagaactggggaagaa 780
Db 721 gtcaagcaagcaactcaagtcacaatgaagcgcgtcactctagaactggggaagaa 780
QY 781 ccaaccatcgtcttcacgaagctctctcgaacggcagtcggggaatcgcgaagat 840
Db 781 ccaaccatcgtcttcacgaagctctctcgaacggcagtcggggaatcgcgaagat 840
QY 841 ttctcaaaattcgggcacaaatttgggtcccccctcctcttctgctagtcgaatgggaat 900
Db 841 ttctcaaaattcgggcacaaatttgggtcccccctcctcttctgctagtcgaatgggaat 900
QY 901 ttgagcgagaaattccatggatccgctcatggtctcatttggagctgtcagagatggctt 960
Db 901 ttgagcgagaaattccatggatccgctcatggtctcatttggagctgtcagagatggctt 960
QY 961 ggcagaaacccattgaaacccaaaggaagcagatggttcccttcgtcgacaagtcacagtac 1020
Db 961 ggcagaaacccattgaaacccaaaggaagcagatggttcccttcgtcgacaagtcacagtac 1020
QY 1021 gacagagcttgggttaacattgacgttggcaaggaatccgcgcagctcctcactggcgtt 1080
Db 1021 gacagagcttgggttaacattgacgttggcaaggaatccgcgcagctcctcactggcgtt 1080
QY 1081 ggtagaaagggcgacaaaggttcgcgattggaacgcagcagatatttgtcaatcccaaaaca 1140
Db 1081 ggtagaaagggcgacaaaggttcgcgattggaacgcagcagatatttgtcaatcccaaaaca 1140
QY 1141 ggcagcaaaatttgggttagagagatcttggcccgcttctgtccattaaagacgttcaag 1200
Db 1141 ggcagcaaaatttgggttagagagatcttggcccgcttctgtccattaaagacgttcaag 1200
QY 1201 acggaagaaagccattgagattgccaatgacagacttatgggctagcctcggtcatt 1260
Db 1201 acggaagaaagccattgagattgccaatgacagacttatgggctagcctcggtcatt 1260
```

```
QY 1261 tatacaaatctctcaaacagggtctccgtgtctcgttcggcgctcagagacgggtggcgctc 1320
Db 1261 tatacaaatctctcaaacagggtctccgtgtctcgttcggcgctcagagacgggtggcgctc 1320
QY 1321 tcgatacaacttccctttatcccgagacacaaactccgtttggcggtatgaacaacatcg 1380
Db 1321 tcgatacaacttccctttatcccgagacacaaactccgtttggcggtatgaacaacatcg 1380
QY 1381 ggtcgaagcagagagtaggcgaagaagggttcgaagggttacttggagcccaagaccatt 1440
Db 1381 ggtcgaagcagagagtaggcgaagaagggttcgaagggttacttggagcccaagaccatt 1440
QY 1441 aatatccagctcaacatagatga 1464
Db 1441 aatatccagctcaacatagatga 1464

RESULT 5
US-09-677-488A-4
; Sequence 4, Application US/09677488A
; GENERAL INFORMATION:
; APPLICANT: Duwick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Craata, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 35718/204100
; CURRENT APPLICATION NUMBER: US/09/677,488A
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: aldehyde dehydrogenase, fully spliced cDNA
US-09-677-488A-4
```

```
Query Match 100.0%; Score 1464; DB 26; Length 1464;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 atgtttcttcgctgacgaatacagaagtgaaactcttcatcaacaatgaattcgtctcc 60
Db 1 atgtttcttcgctgacgaatacagaagtgaaactcttcatcaacaatgaattcgtctcc 60
QY 61 tcaaaaggggttcgagagattaaacgtctcagaaacccgtgggagaaatccacgtttccact 120
Db 61 tcaaaaggggttcgagagattaaacgtctcagaaacccgtgggagaaatccacgtttccact 120
QY 121 gatgttcacgtggccaaacgcgcgcgtatgtcgacagtcgagtagccgttcgttcgagcg 180
Db 121 gatgttcacgtggccaaacgcgcgcgtatgtcgacagtcgagtagccgttcgttcgagcg 180
QY 181 gtcaaaaaggcccatggaagaagttcacaagttcacaagtcgcgtcgttcgagtcgcaatga 240
Db 181 gtcaaaaaggcccatggaagaagttcacaagttcacaagtcgcgtcgttcgagtcgcaatga 240
QY 241 ttgcggagactcccgagagaaacccgcgagaaagctcgtctcgttcgagtcgcaatga 300
Db 241 ttgcggagactcccgagagaaacccgcgagaaagctcgtctcgttcgagtcgcaatga 300
QY 301 ggtagaccggttcgatgatactactatttcacattcccaaatggtctcgttttgcg 360
Db 301 ggtagaccggttcgatgatactactatttcacattcccaaatggtctcgttttgcg 360
```



```

Db 601 gattccctctggaagctgagcttctcactgagacacagtgacgggtgaaagcattg 660
Qy 661 gctgcacatgacattgcaagatcagcttcaagaatctgtcgcggtggcgcgccc 720
Db 661 gctgcacatgacattgcaagatcagcttcaagaatctgtcgcggtggcgcgccc 720
Qy 721 gtaagcaagcaactcaagtcacacatgaagcgcgtcacttagaactgggggaaag 780
Db 721 gtaagcaagcaactcaagtcacacatgaagcgcgtcacttagaactgggggaaag 780
Qy 781 ccaacatcgtctcaacgaactcctctcgaacggcagtcgggggaatcgcaaaagat 840
Db 781 ccaacatcgtctcaacgaactcctctcgaacggcagtcgggggaatcgcaaaagat 840
Qy 841 ttctcaaaattcgggcaaaattgggtcccccctcctctgttgcctagtgcaatgggaaat 900
Db 841 ttctcaaaattcgggcaaaattgggtcccccctcctctgttgcctagtgcaatgggaaat 900
Qy 901 ttacggaagaaattcattgagctcgtcattggtcatttgagagctgcagagatgctt 960
Db 901 ttacggaagaaattcattgagctcgtcattggtcatttgagagctgcagagatgctt 960
Qy 961 ggcagaacccattggaaccccaaggagcagctggtccctctctgcgaacgctccagtac 1020
Db 961 ggcagaacccattggaaccccaaggagcagctggtccctctctgcgaacgctccagtac 1020
Qy 1021 gacagagcttgggttaacattgacgttggaagagatgacgcgcagctcctcaetggcgtt 1080
Db 1021 gacagagcttgggttaacattgacgttggaagagatgacgcgcagctcctcaetggcgtt 1080
Qy 1081 ggtagaagggcgcaagggatcgcgattgaacgcagcagatatttgcatacccaaccca 1140
Db 1081 ggtagaagggcgcaagggatcgcgattgaacgcagcagatatttgcatacccaaccca 1140
Qy 1141 ggcagcaaaattgggttggagagatcttggccctctgttgcatttaagacgttcaag 1200
Db 1141 ggcagcaaaattgggttggagagatcttggccctctgttgcatttaagacgttcaag 1200
Qy 1201 acggaagagagccattgagattcccaatgacacagcattatggctagcctcgctcatt 1260
Db 1201 acggaagagagccattgagattcccaatgacacagcattatggctagcctcgctcatt 1260
Qy 1261 tatcaaaattctcaacagggttccgtgtctcgtcgcgcgtcgcagacgggtggcgtc 1320
Db 1261 tatcaaaattctcaacagggttccgtgtctcgtcgcgcgtcgcagacgggtggcgtc 1320
Qy 1321 tcgatcaacttccctttatcccgagacacaaactccgttggcgcgtatgaacaaatcg 1380
Db 1321 tcgatcaacttccctttatcccgagacacaaactccgttggcgcgtatgaacaaatcg 1380
Qy 1381 ggtcagcgagagctagcggaagagggtcgaaggcgtcacttggaagcccaagaccatt 1440
Db 1381 ggtcagcgagagctagcggaagagggtcgaaggcgtcacttggaagcccaagaccatt 1440
Qy 1441 aatatccagctcaacatagatga 1464
Db 1441 aatatccagctcaacatagatga 1464

```

RESULT 7

US-09-677-682A-4

; Sequence 4, Application US/09677682A

; GENERAL INFORMATION:

; APPLICANT: Duvick, Jon

; APPLICANT: Maddox, Joyce

; APPLICANT: Gilliam, Jacob

; APPLICANT: Folkerts, Otto

; APPLICANT: Crasta, Oswald R.

; TITLE OF INVENTION: Compositions and Methods for Fumonisin

; TITLE OF INVENTION: Detoxification

; FILE REFERENCE: 35718/204101

; CURRENT APPLICATION NUMBER: US/09/677,682A

; CURRENT FILING DATE: 2000-10-02

```

; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: aldehyde dehydrogenase, fully spliced cDNA
US-09-677-682A-4

```

```

Query Match 100.0%; Score 1464; DB 26; Length 1464;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 atggtctcttcgctgacgaatacaagaagtgaactcttcaatacaaatgaattcgtccc 60
Db 1 atggtctcttcgctgacgaatacaagaagtgaactcttcaatacaaatgaattcgtccc 60
Qy 61 tccaaggggttcgagagattaacgctcacgaacccgtggagcaaatccacgttgccat 120
Db 61 tccaaggggttcgagagattaacgctcacgaacccgtggagcaaatccacgttgccat 120
Qy 121 gatgttcacgtggcacaacgcgcgagatgctcagacagtcagtagccgttcggtgcaagg 180
Db 121 gatgttcacgtggcacaacgcgcgagatgctcagacagtcagtagccgttcggtgcaagg 180
Qy 181 gtcaaaagggtcccatggagaagtgtcacagggtgcacaaacgcgcgctgcattgaag 240
Db 181 gtcaaaagggtcccatggagaagtgtcacagggtgcacaaacgcgcgctgcattgaag 240
Qy 241 ttccggacccctcgcgcgagaaacgcgcgagagctcgtctgtgagtcgctgcgtccacc 300
Db 241 ttccggacccctcgcgcgagaaacgcgcgagagctcgtctgtgagtcgctgcgtccacc 300
Qy 301 ggtacacgggtgtcagatcatttcatttcgacattcccaaatgctccgttcttcg 360
Db 301 ggtacacgggtgtcagatcatttcatttcgacattcccaaatgctccgttcttcg 360
Qy 361 tactatgaggtggcgcgacaaagatcgcgcgaaagaccttcccagagacaacgcaag 420
Db 361 tactatgaggtggcgcgacaaagatcgcgcgaaagaccttcccagagacaacgcaag 420
Qy 421 ccgaattggtttacgagccgagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
Db 421 ccgaattggtttacgagccgagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
Qy 481 ttctttacgtcgtcgtggaagatagccccccctcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
Db 481 ttctttacgtcgtcgtggaagatagccccccctcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
Qy 541 aaagctcgtggaataccccctcgtgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 600
Db 541 aaagctcgtggaataccccctcgtgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 600
Qy 601 ggtatccctcgtgagctgtcagcttccctcactgagacacagatgacgggtgaagcattg 660
Db 601 ggtatccctcgtgagctgtcagcttccctcactgagacacagatgacgggtgaagcattg 660
Qy 661 gctgcacatggaattgcaagatcagcttcaagaatcgtcgcggtggcgcgccc 720
Db 661 gctgcacatggaattgcaagatcagcttcaagaatcgtcgcggtggcgcgccc 720
Qy 721 gtcaagcaagcaactcaagtcacacatgaagcgcgtcacttagaactgggggaaag 780
Db 721 gtcaagcaagcaactcaagtcacacatgaagcgcgtcacttagaactgggggaaag 780
Qy 781 ccaacatcgtctcaacgaagctcctctcgaacggcagtcgggggaatcgccaaagggat 840
Db 781 ccaacatcgtctcaacgaagctcctctcgaacggcagtcgggggaatcgccaaagggat 840

```

| | | | |
|----|------|--|------|
| Db | 781 | ccaaccatcgcttccaagaagctcctctcgaacgcgcagtcgggggaatctcgcaagaagat | 840 |
| QY | 841 | tctcaaaattcgggcaaaatttgggtccccccctctgttctgtadtgcatttggggaaaat | 900 |
| Db | 841 | tctcaaaattcgggcaaaatttgggtccccccctctgttctgtadtgcatttggggaaaat | 900 |
| QY | 901 | ttagcggagaattccattggagtcocgtcatggtctcatttggaggtcgtcagagatggcct | 960 |
| Db | 901 | ttagcggagaattccattggagtcocgtcatggtctcatttggaggtcgtcagagatggcct | 960 |
| QY | 961 | ggcgaagaaccttggaaaccaagagacgcatggtccctctcgcacaagtccacgtac | 1020 |
| Db | 961 | ggcgaagaaccttggaaaccaagagacgcatggtccctctcgcacaagtccacgtac | 1020 |
| QY | 1021 | gacagactctgggttaacatttgaacttggcaagatataccgcagactctcactggcct | 1080 |
| Db | 1021 | gacagactctgggttaacatttgaacttggcaagatataccgcagactctcactggcct | 1080 |
| QY | 1081 | ggtagaagaaggcgacaaggattcgcgattgaaccgcagatatttgcattccccaaaca | 1140 |
| Db | 1081 | ggtagaagaaggcgacaaggattcgcgattgaaccgcagatatttgcattccccaaaca | 1140 |
| QY | 1141 | ggcagcaaaatttggttgagagatctttggccccctctgttccatttaagacgttcaag | 1200 |
| Db | 1141 | ggcagcaaaatttggttgagagatctttggccccctctgttccatttaagacgttcaag | 1200 |
| QY | 1201 | acggaagaaggcgcatctagatctgccaattgacacgacttattgggtacgtccgtcatt | 1260 |
| Db | 1201 | acggaagaaggcgcatctagatctgccaattgacacgacttattgggtacgtccgtcatt | 1260 |
| QY | 1261 | tataccaattctcaacagggtctcgtgtctctgttcggcgtctgagaccgttggtc | 1320 |
| Db | 1261 | tataccaattctcaacagggtctcgtgtctctgttcggcgtctgagaccgttggtc | 1320 |
| QY | 1321 | tcgataacttcccctttatcccccgagacacaaactccgttttggcgcatgaacaatcg | 1380 |
| Db | 1321 | tcgataacttcccctttatcccccgagacacaaactccgttttggcgcatgaacaatcg | 1380 |
| QY | 1381 | ggctcaggcagagagcttagcggaagaagggtcgaagcgtacttggagcccaagaccatt | 1440 |
| Db | 1381 | ggctcaggcagagagcttagcggaagaagggtcgaagcgtacttggagcccaagaccatt | 1440 |
| QY | 1441 | aatatccacgtccaacatagagtga | 1464 |
| Db | 1441 | aatatccacgtccaacatagagtga | 1464 |

```

RESULT      8
US-09-677-682A-4
  : Sequence 4, Application US/09677682B
  : GENERAL INFORMATION:
  : APPLICANT: Duvick, Jon
  : APPLICANT: Maddox, Joyce
  : APPLICANT: Gilling, Jacob
  : APPLICANT: Folkerts, Otto
  : APPLICANT: Crasta, Oswald R.
  : TITLE OF INVENTION: Compositions and Methods for Fumonisin
  : TITLE OF INVENTION: Detoxification
  : FILE REFERENCE: 35718/204101
  : CURRENT APPLICATION NUMBER: US/09/677,682B
  : CURRENT FILING DATE: 2000-10-02
  : PRIOR APPLICATION NUMBER: 09/355,224
  : PRIOR FILING DATE: 1999-07-12
  : NUMBER OF SEQ ID NOS: 11
  : SOFTWARE: FASTSEQ for Windows Version 4.0
  : SEQ ID NO 4
  : LENGTH: 1464
  : TYPE: DNA
  : ORGANISM: Exophiala spinifera
  : FEATURE:
  : NAME/KEY: misc_feature
  : LOCATION: (0)...(0)
  : OTHER INFORMATION: aldephde dehydrogenase, fully spliced

```

US-09-677-682B-4
 Query Match 100.00; Score 1464; DB 26; Length 1464;
 Best Local Similarity 100.00; Pred. No. 0;
 Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|---|------|
| Qy | 1 | atggtttcttcgctgcagcaatacaagatgaactcttcatacaaatgaattcgctccc | 60 |
| Db | 1 | atggtttcttcgctgcagcaatacaagatgaactcttcatacaaatgaattcgctccc | 60 |
| Qy | 61 | tccaaagggctcgagagattaaagctcaagaccctggagagaaatccacagcttcccact | 120 |
| Db | 61 | tccaaagggctcgagagattaaagctcaagaccctggagagaaatccacagcttcccact | 120 |
| Qy | 121 | gattgttcacgttgcccaacgcgcgcagatgtcgacagtacagtagccgtcttcggtgcagcgc | 180 |
| Db | 121 | gattgttcacgttgcccaacgcgcgcagatgtcgacagtacagtagccgtcttcggtgcagcgc | 180 |
| Qy | 181 | gtcaaaaagggcccatggaaagagttcacagggtcacaaagcgcgcgcgtgcagcttaag | 240 |
| Db | 181 | gtcaaaaagggcccatggaaagagttcacagggtcacaaagcgcgcgcgtgcagcttaag | 240 |
| Qy | 241 | ttcgcggaaacctgcgcgaagaagaacgcgcgaagaagctgcgtctgtgagtcgctgcacacc | 300 |
| Db | 241 | ttcgcggaaacctgcgcgaagaagaacgcgcgaagaagctgcgtctgtgagtcgctgcacacc | 300 |
| Qy | 301 | ggtagaccggtgtcgatgatcacctatttcgacattccaaacatggctcccggttttcgc | 360 |
| Db | 301 | ggtagaccggtgtcgatgatcacctatttcgacattccaaacatggctcccggttttcgc | 360 |
| Qy | 361 | tactatgcaggtgggcgcgaagaatgcgcgaagaacacatttcgcgaagacaacggaag | 420 |
| Db | 361 | tactatgcaggtgggcgcgaagaatgcgcgaagaacacatttcgcgaagacaacggaag | 420 |
| Qy | 421 | ccgaattggcgcttacagaccatgggggtgtgctgctgattatggccaagctgcgaacgcgact | 480 |
| Db | 421 | ccgaattggcgcttacagaccatgggggtgtgctgctgattatggccaagctgcgaacgcgact | 480 |
| Qy | 481 | ttctttacgttcggttcgaaagatagccccgcctgcgcgcgcgtgccttcattcttc | 540 |
| Db | 481 | ttctttacgttcggttcgaaagatagccccgcctgcgcgcgcgtgccttcattcttc | 540 |
| Qy | 541 | aaagctcgcgaataatccccctggtggatctctgggctgcctctcttcgcagaagcc | 600 |
| Db | 541 | aaagctcgcgaataatccccctggtggatctctgggctgcctctcttcgcagaagcc | 600 |
| Qy | 601 | ggattcccccttcgtgagtcgtagtctacttcactggagcagcagtgacgggtgaagcattg | 660 |
| Db | 601 | ggattcccccttcgtgagtcgtagtctacttcactggagcagcagtgacgggtgaagcattg | 660 |
| Qy | 661 | gcgtgcacatgggaacattgcgaagaatcagcttcacaagaatctctgcgggtggccgcgcc | 720 |
| Db | 661 | gcgtgcacatgggaacattgcgaagaatcagcttcacaagaatctctgcgggtggccgcgcc | 720 |
| Qy | 721 | gtcaagcagcgaacactcaagtccaaacatgaagcgcgtcactctagaactcgggggaaaaag | 780 |
| Db | 721 | gtcaagcagcgaacactcaagtccaaacatgaagcgcgtcactctagaactcgggggaaaaag | 780 |
| Qy | 781 | ccaacacatcgtcttcaacgaagctcctctcgaaacgcgcagtcgggggaaatcgcgcaaggaat | 840 |
| Db | 781 | ccaacacatcgtcttcaacgaagctcctctcgaaacgcgcagtcgggggaaatcgcgcaaggaat | 840 |
| Qy | 841 | ttctcaaaattcgggcaaaatttgggtccccccctctctgtctcagtgcgaatggggagaaat | 900 |
| Db | 841 | ttctcaaaattcgggcaaaatttgggtccccccctctctgtctcagtgcgaatggggagaaat | 900 |
| Qy | 901 | ttagcggagaaatttccattggagtcgctcatgggtctatcttggaggctctcagagatcggtt | 960 |
| Db | 901 | ttagcggagaaatttccattggagtcgctcatgggtctatcttggaggctctcagagatcggtt | 960 |
| Qy | 961 | ggccagaacccattggaaacccaaagagacgcagtggttcccttcgtcgaacagtcaccagttac | 1020 |

Db 961 ggcagaaacccattggaaccacaaagagacgacgtggtcccttcgtcgacaagtccacgtac 1020
Qy 1021 gacagagttctgggttaacattgacgttggcaagagataccgcgacgtctctcactggcggt 1080
Db 1021 gacagagttctgggttaacattgacgttggcaagagataccgcgacgtctctcactggcggt 1080
Qy 1081 ggtagaagaggcgacaaaggattccgcgattgaaccgacgatatattgtcaatcccaaacca 1140
Db 1081 ggtagaagaggcgacaaaggattccgcgattgaaccgacgatatattgtcaatcccaaacca 1140
Qy 1141 ggcagcaaaatttgggttagagagattttggcccgcttctccattaaagacgttcaag 1200
Db 1141 ggcagcaaaatttgggttagagagattttggcccgcttctccattaaagacgttcaag 1200
Qy 1201 acggaagaaggccattgagattgccaatgacacgacttatgggtagacctgggtcatt 1260
Db 1201 acggaagaaggccattgagattgccaatgacacgacttatgggtagacctgggtcatt 1260
Qy 1261 tataccaaatctcaacagagggtctcggtgctcgtcgcgctcgagaccggtggcgctc 1320
Db 1261 tataccaaatctcaacagagggtctcggtgctcgtcgcgctcgagaccggtggcgctc 1320
Qy 1321 tcgatacaattccctttatcccccgagacacaaactccgtttggcgcatgaaacaatcg 1380
Db 1321 tcgatacaattccctttatcccccgagacacaaactccgtttggcgcatgaaacaatcg 1380
Qy 1381 ggcctcaggcagagagctaggcgaaagggtctcaaggcgtacttggagcccaagaccatt 1440
Db 1381 ggcctcaggcagagagctaggcgaaagggtctcaaggcgtacttggagcccaagaccatt 1440
Qy 1441 aatatccacgtcaacatagatga 1464
Db 1441 aatatccacgtcaacatagatga 1464

RESULT 9

US-09-882-694-4
; Sequence 4, Application US/09882694
; GENERAL INFORMATION:
; APPLICANT: DuVick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 35718/208255
; CURRENT APPLICATION NUMBER: US/09/882,694
; PRIOR FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: aldehyde dehydrogenase, fully spliced cDNA
US-09-882-694-4

Query Match 100.0%; Score 1464; DB 33; Length 1464;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 atgggtttcttcgctgacgaatacaagagtgaactttctcaatcaacaatgaattcgctccc 60
Db 1 atgggtttcttcgctgacgaatacaagagtgaactttctcaatcaacaatgaattcgctccc 60
Qy 61 tccaaggggtccgagagattaacgtcacgaaccggtggagcaatccacgttgccact 120

Db 61 tccaaggggtccgagagattaacgtcacgaaccggtggagcaatccacgttgccact 120
Qy 121 gatgttcaactggccaaacgcgcgcatgtcgacagtgcagttagccgcttcggtgcagcg 180
Db 121 gatgttcaactggccaaacgcgcgcatgtcgacagtgcagttagccgcttcggtgcagcg 180
Qy 181 gtcacaaagggcccatgggaagaattcacaggtgcacaaacgcgcgctgcattcctaaag 240
Db 181 gtcacaaagggcccatgggaagaattcacaggtgcacaaacgcgcgctgcattcctaaag 240
Qy 241 ttgcggagacctcccgagaagaacgcgcgagagctcgctcgtcgtgaactcgtcccaac 300
Db 241 ttgcggagacctcccgagaagaacgcgcgagagctcgctcgtcgtgaactcgtcccaac 300
Qy 301 ggtagaccggtgctgatgatcactcatttcgcacattccaaacatggtctcgtgtttcgc 360
Db 301 ggtagaccggtgctgatgatcactcatttcgcacattccaaacatggtctcgtgtttcgc 360
Qy 361 tactatgcagctggccgcacaaagatcgcgaaagacatttcccagagacaacggcaag 420
Db 361 tactatgcagctggccgcacaaagatcgcgaaagacatttcccagagacaacggcaag 420
Qy 421 ccgaattggcgttacgagccgattgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
Db 421 ccgaattggcgttacgagccgattgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
Qy 481 ttcttttaactgcgctgggaagatagccccgcctcgcgcgcgcgcgcgcgcgcgcgcgc 540
Db 481 ttcttttaactgcgctgggaagatagccccgcctcgcgcgcgcgcgcgcgcgcgcgcgc 540
Qy 541 aaagcctcggagaaatcccccgtggcggttctgggctcgcctcctctctcctcctcctc 600
Db 541 aaagcctcggagaaatcccccgtggcggttctgggctcgcctcctctcctcctcctcct 600
Qy 601 ggattccctcctggagtcgtgcagttcctcactggagcacgagtgacgggtgaaacattg 660
Db 601 ggattccctcctggagtcgtgcagttcctcactggagcacgagtgacgggtgaaacattg 660
Qy 661 gctgcacatgacattgcgaagatcagcttcacaaagatctcgcgcgcgcgcgcgcgcgc 720
Db 661 gctgcacatgacattgcgaagatcagcttcacaaagatctcgcgcgcgcgcgcgcgcgc 720
Qy 721 gtaacgaacacactcaagtccaacatgaagcgcgtcactctagaaactgggggaaag 780
Db 721 gtaacgaacacactcaagtccaacatgaagcgcgtcactctagaaactgggggaaag 780
Qy 781 ccaaccatcgtcttcacgaagctcctcgcgaacgacgtcggggaatcggaacagat 840
Db 781 ccaaccatcgtcttcacgaagctcctcgcgaacgacgtcggggaatcggaacagat 840
Qy 841 ttctcaaaattcgggcaaaattgggttccccccctcctcgtttctagtagtcaatgggaaat 900
Db 841 ttctcaaaattcgggcaaaattgggttccccccctcctcgtttctagtagtcaatgggaaat 900
Qy 901 ttacggagaaaattccatggagtccctcactcatttggggggtcgtcagagatggcgtt 960
Db 901 ttacggagaaaattccatggagtccctcactcatttggggggtcgtcagagatggcgtt 960
Qy 961 ggcagagaccattggaaacaaagagacgattgctcctcgtcgaagatcccaagtac 1020
Db 961 ggcagagaccattggaaacaaagagacgattgctcctcgtcgaagatcccaagtac 1020
Qy 1021 gacagagctcttggttaacattgacgttggcaaggaataccgcgacgtctcctcactggcgt 1080
Db 1021 gacagagctcttggttaacattgacgttggcaaggaataccgcgacgtctcctcactggcgt 1080
Qy 1081 ggtagaagaggcgacaaaggattccgcgattgaaccgacgatatattgtcaatcccaaacca 1140
Db 1081 ggtagaagaggcgacaaaggattccgcgattgaaccgacgatatattgtcaatcccaaacca 1140
Qy 1141 ggcagcaaaatttgggttagagagattttggcccgcttctccattaaagacgttcaag 1200

| | | | |
|----|------|---|------|
| Db | 1141 | ggcgcgaataattggtttgaggagatcttttgcccgctcttgcattaaagacgttcaag | 1200 |
| Oy | 1201 | acggaagaaggccattgagattgccaatgacacgacttatggctagcctcggtcatt | 1260 |
| Db | 1201 | acggaagaaggccattgagattgccaatgacacgacttatggctagcctcggtcatt | 1260 |
| Oy | 1261 | tatacaaatctctcaacagggtctcggctcgtcgcgcgctcgagacccggtggcgctc | 1320 |
| Db | 1261 | tatacaaatctctcaacagggtctcggctcgtcgcgcgctcgagacccggtggcgctc | 1320 |
| Oy | 1321 | tcgatcaacttccctttatccccgagacacaaactcggttggcggtatgaacaactcg | 1380 |
| Db | 1321 | tcgatcaacttccctttatccccgagacacaaactcggttggcggtatgaacaactcg | 1380 |
| Oy | 1381 | ggctcaggcgagagcttagcgcaagaaggctcgaaggcgctacttggagcccaagaccatt | 1440 |
| Db | 1381 | ggctcaggcgagagcttagcgcaagaaggctcgaaggcgctacttggagcccaagaccatt | 1440 |
| Oy | 1441 | aatatccacgtcaacatagatgga | 1464 |
| Db | 1441 | aatatccacgtcaacatagatgga | 1464 |

RESULT 10

```

US-09-882-694A-4
; Sequence 4, Application US/09882694A
; GENERAL INFORMATION:
; APPLICANT: Duwick, Jon
; APPLICANT: Madcox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/882,694A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: aldehyde dehydrogenase, fully spliced cDNA
US-09-882-694A-4

```

```
Query Match      100.0%; Score 1464; DB 33; Length 1464;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

| | | | |
|----|-----|--|-----|
| Qy | 1 | atggttcttctgcctgacgaatacaagagtgaaactctcatacaaatgaattcgtctcc | 60 |
| | | | |
| Db | 1 | atggttcttctgcctgacgaatacaagagtgaaactctcatacaaatgaattcgtctcc | 60 |
| | | | |
| Qy | 61 | tccaagggttcgagaagattaaacgtcacgaaccgtggagcgaatcaccggtgcacct | 120 |
| | | | |
| Db | 61 | tccaagggttcgagaagattaaacgtcacgaaccgtggagcgaatcaccggtgcacct | 120 |
| | | | |
| Qy | 121 | gatgttcacgtggccaacgcggccgagtgcacagtcgagtagccgtctcgggtgcaggcg | 180 |
| | | | |
| Db | 121 | gatgttcacgtggccaacgcggccgagtgcacagtcgagtagccgtctcgggtgcaggcg | 180 |
| | | | |
| Qy | 181 | gtcaaaaaggcccatggagaagttcacagtgacaaacgcgcgctgcattcgttaag | 240 |
| | | | |
| Db | 181 | gtcaaaaaggcccatggagaagttcacagtgacaaacgcgcgctgcattcgttaag | 240 |
| | | | |
| Qy | 241 | ttcgcggacctctgcgagagaagacgcgcgaagctcgtctctcggaagtcgctgccacc | 300 |


```

; APPLICANT: Hammond-Kosack, Kim
; APPLICANT: Masucci, James D.
; APPLICANT: Urban, Martin
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE OF INVENTION: PLANTS
; FILE REFERENCE: 38-21(51838)A
; CURRENT APPLICATION NUMBER: US/60/189,657
; CURRENT FILING DATE: 2000-03-15
; NUMBER OF SEQ ID NOS: 22822
; SEQ ID NO 13072
; LENGTH: 340
; TYPE: DNA
; ORGANISM: Triticum aestivum
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3399-009-P1-K1-B6
US-60-189-657-13072

```

```

Query Match      1.4%; Score 21; DB 57; Length 340;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 502 atagccccgcctgcgcgcc 522
    |||||
Db 307 atagccccgcctgcgcgcc 327

```

Search completed: April 27, 2002, 07:42:26
Job time: 21774 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 05:54:18 ; Search time 151.01 Seconds
(without alignments)
2381.349 Million cell updates/sec

Title: US-09-882-694A-4
Perfect score: 1464
Sequence: 1 atgggtcttgcgtgacga.....tccacgtcaacatagagtga 1464

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 123816752 residues

Word size : 0
Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database : Issued Patents, NA.*
- 1: /cgn2_6/ptodata/2/ina/5A-COMB.seq.*
 - 2: /cgn2_6/ptodata/2/ina/5B-COMB.seq.*
 - 3: /cgn2_6/ptodata/2/ina/6A-COMB.seq.*
 - 4: /cgn2_6/ptodata/2/ina/6B-COMB.seq.*
 - 5: /cgn2_6/ptodata/2/ina/ECTUS-COMB.seq.*
 - 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------------|
| c 1 | 18 | 1.2 | 1125 | 4 | US-09-516-914-12 |
| 2 | 18 | 1.2 | 1262 | 2 | US-08-989-386-4 |
| 3 | 18 | 1.2 | 1839 | 1 | US-08-442-248-3 |
| 4 | 18 | 1.2 | 1839 | 1 | US-08-440-815-3 |
| 5 | 18 | 1.2 | 1839 | 4 | US-08-486-449-3 |
| 6 | 18 | 1.2 | 14272 | 4 | US-09-516-914-23 |
| 7 | 17 | 1.2 | 111 | 4 | US-09-127-680-11 |
| 8 | 17 | 1.2 | 340 | 3 | US-08-441-971-10 |
| 9 | 17 | 1.2 | 340 | 3 | US-08-441-971-11 |
| 10 | 17 | 1.2 | 340 | 4 | US-08-221-653-10 |
| 11 | 17 | 1.2 | 340 | 4 | US-08-221-653-11 |
| 12 | 17 | 1.2 | 340 | 4 | US-08-442-144A-10 |
| 13 | 17 | 1.2 | 340 | 4 | US-08-442-144A-11 |
| 14 | 17 | 1.2 | 340 | 4 | US-08-441-970-10 |
| 15 | 17 | 1.2 | 340 | 4 | US-08-441-970-11 |
| 16 | 17 | 1.2 | 1079 | 1 | US-08-471-570-13 |
| 17 | 17 | 1.2 | 1950 | 2 | US-08-377-440A-2 |
| c 18 | 17 | 1.2 | 1950 | 4 | US-09-440-530-2 |
| 19 | 17 | 1.2 | 1954 | 1 | US-08-471-570-5 |
| 20 | 17 | 1.2 | 2181 | 4 | US-09-158-767-10 |
| 21 | 17 | 1.2 | 2676 | 1 | US-08-471-570-7 |
| 22 | 17 | 1.2 | 2991 | 1 | US-08-324-977-49 |
| 23 | 17 | 1.2 | 2991 | 2 | US-08-384-616-49 |
| 24 | 17 | 1.2 | 2991 | 2 | US-08-904-686A-49 |
| 25 | 17 | 1.2 | 2991 | 4 | US-09-315-850-49 |
| 26 | 17 | 1.2 | 3416 | 2 | US-08-451-822A-15 |
| 27 | 17 | 1.2 | 3416 | 4 | US-08-323-430-15 |

| | | | | | | |
|----|----|-----|------|---|-------------------|-------------------|
| 28 | 17 | 1.2 | 7001 | 1 | US-08-258-261B-1 | Sequence 1, Appli |
| 29 | 17 | 1.2 | 7001 | 1 | US-08-456-837-1 | Sequence 1, Appli |
| 30 | 17 | 1.2 | 7001 | 1 | US-08-457-342-1 | Sequence 1, Appli |
| 31 | 17 | 1.2 | 7001 | 1 | US-08-457-646A-1 | Sequence 1, Appli |
| 32 | 17 | 1.2 | 7001 | 1 | US-08-458-076A-1 | Sequence 1, Appli |
| 33 | 17 | 1.2 | 7001 | 1 | US-08-457-335A-1 | Sequence 1, Appli |
| 34 | 17 | 1.2 | 7001 | 1 | US-08-761-258-6 | Sequence 6, Appli |
| 35 | 17 | 1.2 | 7001 | 1 | US-08-729-214-1 | Sequence 1, Appli |
| 36 | 17 | 1.2 | 7001 | 2 | US-08-977-306-6 | Sequence 6, Appli |
| 37 | 17 | 1.2 | 7001 | 3 | US-03-028-934-1 | Sequence 1, Appli |
| 38 | 17 | 1.2 | 7863 | 1 | US-08-324-977-35 | Sequence 35, Appl |
| 39 | 17 | 1.2 | 7863 | 2 | US-08-384-616-35 | Sequence 35, Appl |
| 40 | 17 | 1.2 | 7863 | 2 | US-08-904-686A-35 | Sequence 35, Appl |
| 41 | 17 | 1.2 | 7863 | 4 | US-09-315-850-35 | Sequence 35, Appl |
| 42 | 17 | 1.2 | 7917 | 1 | US-08-324-977-31 | Sequence 31, Appl |
| 43 | 17 | 1.2 | 7917 | 2 | US-08-384-616-31 | Sequence 31, Appl |
| 44 | 17 | 1.2 | 7917 | 2 | US-08-904-686A-31 | Sequence 31, Appl |
| 45 | 17 | 1.2 | 7917 | 4 | US-09-315-850-31 | Sequence 31, Appl |

ALIGNMENTS

RESULT 1
US-09-516-914-12/c
; Sequence 12, Application US/09516914
; Patent No. 6333401
; GENERAL INFORMATION:
; APPLICANT: Breinig, Sabine
; APPLICANT: Fuchs, Georg
; TITLE OF INVENTION: Phenol-Induced Proteins of Thauera aromatica
; FILE REFERENCE: BC1006 US NA
; CURRENT APPLICATION NUMBER: US/09/516,914
; EARLIER FILING DATE: 2000-03-01
; EARLIER APPLICATION NUMBER: 60/122,952
; EARLIER FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 12
; LENGTH: 1125
; TYPE: DNA
; ORGANISM: Thauera aromatica
US-09-516-914-12

Query Match 1.2%; Score 18; DB 4; Length 1125;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 701 ctctcgcgcgtgcccgcg 718
|||||
Db 50 CTGTCGGCGGTGCGCGCG 33

RESULT 2
US-08-989-386-4
; Sequence 4, Application US/08989386
; Patent No. 5989860
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN ISOMERASE HOMOLOGS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/989,386
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0443 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1262 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TMLRDT01
; CLONE: 292808
; US-08-989-386-4

Query Match 1.2% Score 18; DB 2; Length 1262;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 767 aactgggggaaagccaa 784
Db 956 AACTGGGGGAAAGCCAA 973

RESULT 3
; Sequence 3, Application US/08442248
; Patent No. 5759863
; GENERAL INFORMATION:
; APPLICANT: Caras, Ingrid W.
; TITLE OF INVENTION: AL-1 Neurotrophic Factor
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/442,248
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/330128
; FILING DATE: 27-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Torchia, Timothy E.
; REGISTRATION NUMBER: 36,700
; REFERENCE/DOCKET NUMBER: 920C4

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-8674
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1839 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-442-248-3

Query Match 1.2% Score 18; DB 1; Length 1839;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 506 ccccgccctcgccgcg 523
Db 6 CCCCCCGCCCTCGCGCG 23

RESULT 4
; US-08-440-815-3
; Sequence 3, Application US/08440815
; Patent No. 5798448
; GENERAL INFORMATION:
; APPLICANT: Caras, Ingrid W.
; TITLE OF INVENTION: AL-1 Neurotrophic Factor
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,815
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/330128
; FILING DATE: 27-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Torchia, Timothy E.
; REGISTRATION NUMBER: 36,700
; REFERENCE/DOCKET NUMBER: 920C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-8674
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1839 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-440-815-3

Query Match 1.2% Score 18; DB 1; Length 1839;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 506 ccccgccctcgccgcg 523
Db 6 CCCCCCGCCCTCGCGCG 23

```
Db 6 CCCCCGCCCTGGCGCGCG 23

RESULT 5
US-08-486-449-3
; Sequence 3, Application US/08486449
; Patent No. 6280732
; GENERAL INFORMATION:
; APPLICANT: Caras, Ingrid W.
; APPLICANT: Winslow, John W.
; TITLE OF INVENTION: AL-1 Neurotrophic Factor
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 MB floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,449
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/330128
; FILING DATE: 27-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Torchia, Timothy E.
; REGISTRATION NUMBER: 36,700
; REFERENCE/DOCKET NUMBER: P0920P1
; TELEPHONE: 415/225-8674
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; LENGTH: 1839 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-486-449-3

Query Match 1.2%; Score 18; DB 4; Length 1839;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 506 cccccgccctggcgcgcg 523
Db 6 CCCCCGCCCTGGCGCGCG 23

RESULT 6
US-09-516-914-23/c
; Sequence 23, Application US/09516914
; Patent No. 6333401
; GENERAL INFORMATION:
; APPLICANT: Breinig, Sabine
; APPLICANT: Fuchs, Georg
; TITLE OF INVENTION: Phenol-induced Proteins of Thauera aromatica
; FILE REFERENCE: ECI006 US NA
; CURRENT APPLICATION NUMBER: US/09/516,914
; CURRENT FILING DATE: 2000-03-01
; EARLIER APPLICATION NUMBER: 60/122,952
; EARLIER FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 23

; LENGTH: 14272
; TYPE: DNA
; ORGANISM: Thauera aromatica
US-09-516-914-23

Query Match 1.2%; Score 18; DB 4; Length 14272;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 701 ctgtcgcggtggcgcg 718
Db 4765 CTGTGCGGTGGCGCG 4748

RESULT 7
US-09-127-680-11
; Sequence 11, Application US/09127680A
; Patent No. 6235973
; GENERAL INFORMATION:
; APPLICANT: Sanford, John
; APPLICANT: Blowers, Alan
; APPLICANT: Smith, Franzine
; APPLICANT: Vaneck, Joyce
; APPLICANT: Sanford Scientific, Inc.
; TITLE OF INVENTION: Expression of Antimicrobial Peptide Genes in Plants,
; and Their Use in Creating Resistance to Multiple Plant
; TITLE OF INVENTION: Pathogens
; FILE REFERENCE: Antimicrobial peptides
; CURRENT APPLICATION NUMBER: US/09/127,680A
; CURRENT FILING DATE: 1998-07-31
; EARLIER APPLICATION NUMBER: 60/054,315
; EARLIER FILING DATE: 1997-07-31
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 111
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Gene coding
; OTHER INFORMATION: for the extracellular localized antimicrobial
; OTHER INFORMATION: MSI-55 peptide
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(111)
US-09-127-680-11

Query Match 1.2%; Score 17; DB 4; Length 111;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 381 caagatcgcggaaga 397
Db 45 caagatcgcggaaga 61

RESULT 8
US-08-441-971-10
; Sequence 10, Application US/08441971
; Patent No. 6071693
; GENERAL INFORMATION:
; APPLICANT: Tai-An Cha
; TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
; DIAGNOSTICS AND THERAPEUTICS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
```

ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS Version 3.3
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/441,971
FILING DATE: 16-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/221,653
FILING DATE:
APPLICATION NUMBER: US/07/881,528
FILING DATE:
APPLICATION NUMBER: 07/697,326
FILING DATE: 8 May 1991
ATTORNEY/AGENT INFORMATION:
NAME: Janluk, Anthony J.
REGISTRATION NUMBER: 29,809
REFERENCE/DOCKET NUMBER: C0772/7000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 720-3500
TELEFAX: (617) 720-2441
TELEX: EZEKIEL
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 340 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: ns5gh6
US-08-441-971-10

Query Match 1.28; Score 17; DB 3; Length 340;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 975 ggaacccaagagcgc 991
|||||
DB 308 GGAACCCAAGAGGAGC 324

RESULT 9
US-08-441-971-11
Sequence 11, Application US/08441971
Patent No. 6071693
GENERAL INFORMATION:
APPLICANT: Tai-An Cha
TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS Version 3.3
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/441,971
FILING DATE: 16-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/221,653

FILING DATE:
APPLICATION NUMBER: US/07/881,528
FILING DATE:
APPLICATION NUMBER: 07/697,326
FILING DATE: 8 May 1991
ATTORNEY/AGENT INFORMATION:
NAME: Janluk, Anthony J.
REGISTRATION NUMBER: 29,809
REFERENCE/DOCKET NUMBER: C0772/7000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 720-3500
TELEFAX: (617) 720-2441
TELEX: EZEKIEL
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 340 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: ns5spl
US-08-441-971-11

Query Match 1.28; Score 17; DB 3; Length 340;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 975 ggaacccaagagcgc 991
|||||
DB 308 GGAACCCAAGAGGAGC 324

RESULT 10
US-08-221-653-10
Sequence 10, Application US/08221653
Patent No. 6190864
GENERAL INFORMATION:
APPLICANT: Tai-An Cha
TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS Version 3.3
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/221,653
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/881,528
FILING DATE:
APPLICATION NUMBER: 07/697,326
FILING DATE: 8 May 1991
ATTORNEY/AGENT INFORMATION:
NAME: Janluk, Anthony J.
REGISTRATION NUMBER: 29,809
REFERENCE/DOCKET NUMBER: C0772/7000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 720-3500
TELEFAX: (617) 720-2441
TELEX: EZEKIEL
INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:
 LENGTH: 340 nucleotides
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA
 ORIGINAL SOURCE:
 INDIVIDUAL ISOLATE: ns5gh6
 US-08-221-653-10

Query Match 1.2%; Score 17; DB 4; Length 340;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 975 ggaacccaagagacgc 991
 Db 308 GGAACCCAGAGGACGC 324

RESULT 11
 US-08-221-653-11
 ; Sequence 11, Application US/08221653
 ; Patent No. 6190864
 ; GENERAL INFORMATION:
 ; APPLICANT: Tai-An Cha
 ; TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
 ; NUMBER OF SEQUENCES: 147
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
 ; STREET: 600 Atlantic Avenue
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02210
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 5.25 inch
 ; COMPUTER: IBM compatible
 ; OPERATING SYSTEM: MS-DOS version 3.3
 ; SOFTWARE: Wordperfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/221.653
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/881.528
 ; FILING DATE:
 ; FILING DATE: 07/697.326
 ; FILING DATE: 8 May 1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Jantluk, Anthony J.
 ; REGISTRATION NUMBER: 29,809
 ; REFERENCE/DOCKET NUMBER: C0772/7000
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 720-3500
 ; TELEFAX: (617) 720-2441
 ; TELEX: EZEKIEL
 ; INFORMATION FOR SEQ ID NO: 11:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 340 nucleotides
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA
 ; ORIGINAL SOURCE:
 ; INDIVIDUAL ISOLATE: ns5spl
 ; US-08-221-653-11

Query Match 1.2%; Score 17; DB 4; Length 340;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 975 ggaacccaagagacgc 991
 Db 308 GGAACCCAGAGGACGC 324

RESULT 12
 US-08-442-144A-10
 ; Sequence 10, Application US/08442144A
 ; Patent No. 6214583
 ; GENERAL INFORMATION:
 ; APPLICANT: Tai-An Cha
 ; APPLICANT: Eileen Beall
 ; APPLICANT: Bruce Irvine
 ; APPLICANT: Janice Kolberg
 ; APPLICANT: Michael S. Urdea
 ; TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
 ; NUMBER OF SEQUENCES: 148
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Chiron Corporation
 ; STREET: 4560 Horton Street
 ; CITY: Emeryville
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94608-2916
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.5 Inch
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: Windows NT
 ; SOFTWARE: Microsoft Word 97
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/442.144A
 ; FILING DATE: MAY 16, 1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/221.653
 ; FILING DATE: APRIL 1, 1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Doreen Yafko Trujillo
 ; REGISTRATION NUMBER: 35,719
 ; REFERENCE/DOCKET NUMBER: CHIR-0121
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 215-568-3100
 ; TELEFAX: 215-568-3439
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 340 Nucleotides
 ; TYPE: Nucleic Acid
 ; STRANDEDNESS: Single
 ; TOPOLOGY: Linear
 ; MOLECULE TYPE: DNA
 ; ORIGINAL SOURCE:
 ; INDIVIDUAL ISOLATE: ns5gh6
 ; US-08-442-144A-10

Query Match 1.2%; Score 17; DB 4; Length 340;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 975 ggaacccaagagacgc 991
 Db 308 GGAACCCAGAGGACGC 324

RESULT 13
 US-08-442-144A-11
 ; Sequence 11, Application US/08442144A
 ; Patent No. 6214583
 ; GENERAL INFORMATION:
 ; APPLICANT: Tai-An Cha

Query Match 1.2%; Score 17; DB 4; Length 340;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 975 ggaacccaagagacgc 991
 Db 308 GGAACCCAGAGGACGC 324

```
; APPLICANT: Eileen Beall
; APPLICANT: Bruce Irvine
; APPLICANT: Janice Kolberg
; APPLICANT: Michael S. Urdea
; TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 Inch
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows NT
; SOFTWARE: Microsoft Word 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/442,144A
; FILING DATE: MAY 16, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/221,653
; FILING DATE: APRIL 1, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Doreen Yanko Trujillo
; REGISTRATION NUMBER: 35,719
; REFERENCE/DOCKET NUMBER: CHR-0121
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; TELEX:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 Nucleotides
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: DNA
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: ns5spl
; US-08-442-144A-11

Query Match 1.2% Score 17; DB 4; Length 340;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 975 ggaacccaagagagcgc 991
Db 308 GGAACCCAAAGAGACGC 324

RESULT 14
US-08-441-970-10
; Sequence 10, Application US/08441970
; Patent No. 6297370
; GENERAL INFORMATION:
; APPLICANT: Tai-An Cha
; TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 Inch
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,970
; FILING DATE: 16-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/881,528
; FILING DATE: 08-MAY-1992
; APPLICATION NUMBER: 07/697,326
; FILING DATE: 8 MAY 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Janluk, Anthony J.
```

```
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,970
; FILING DATE: 16-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/881,528
; FILING DATE: 08-MAY-1992
; APPLICATION NUMBER: 07/697,326
; FILING DATE: 8 MAY 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Janluk, Anthony J.
; REGISTRATION NUMBER: 29,809
; REFERENCE/DOCKET NUMBER: C0772/7000
; TELEPHONE: (617) 720-3500
; TELEFAX: (617) 720-2441
; TELEX: EZEKIEL
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: ns5gh6
; US-08-441-970-10

Query Match 1.2% Score 17; DB 4; Length 340;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 975 ggaacccaagagagcgc 991
Db 308 GGAACCCAAAGAGACGC 324

RESULT 15
US-08-441-970-11
; Sequence 11, Application US/08441970
; Patent No. 6297370
; GENERAL INFORMATION:
; APPLICANT: Tai-An Cha
; TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 Inch
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,970
; FILING DATE: 16-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/881,528
; FILING DATE: 08-MAY-1992
; APPLICATION NUMBER: 07/697,326
; FILING DATE: 8 MAY 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Janluk, Anthony J.
```

```

; REGISTRATION NUMBER: 29,809
; REFERENCE/DOCKET NUMBER: C0772/7000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 720-3500
; TELEFAX: (617) 720-2441
; TELEX: EZEKTEL
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: ns5spl
; US-08-441-970-11

```

```

Query Match      1.2%; Score 17; DB 4; Length 340;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 975 ggaacccaagagagacgc 991
    |||||
Db 308 GGAACCCAGAGAGACGC 324

```

```

Search completed: April 27, 2002, 05:54:29
Job time: 15356 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 05:28:04 ; Search time 705 Seconds
(without alignments)
3565.337 Million cell updates/sec

Title: US-09-882-694A-4
Perfect score: 1464
Sequence: 1 atggtcttctgcctgacga.....tccacgtcaacatagagtga 1464

Scoring table: OLIGO.NUC
Gapop 60.0., Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 1464 | 100.0 | 1464 | 22 | AA02692 |
| 2 | 20 | 1.4 | 1980 | 23 | ABL10879 |
| 3 | 20 | 1.4 | 3278 | 23 | ABL24684 |
| 4 | 20 | 1.4 | 6153 | 23 | ABL10878 |
| 5 | 19 | 1.3 | 286 | 21 | AA03035 |
| 6 | 19 | 1.3 | 578 | 22 | AAF68055 |
| 7 | 19 | 1.3 | 600 | 22 | AAF22587 |
| 8 | 19 | 1.3 | 828 | 23 | ABL17815 |
| 9 | 19 | 1.3 | 1015 | 22 | AAF22491 |
| | | | | | Exophiala spinifera |
| | | | | | Drosophila melanog |
| | | | | | Drosophila melanog |
| | | | | | Drosophila melanog |
| | | | | | Human secreted pro |
| | | | | | Corynebacterium gl |
| | | | | | Human breast cance |
| | | | | | Drosophila melanog |
| | | | | | Human breast cance |

| | | | | | |
|----|----|-----|-------|----|----------|
| 10 | 19 | 1.3 | 1887 | 22 | AAH67550 |
| 11 | 19 | 1.3 | 2010 | 22 | AAF68054 |
| 12 | 19 | 1.3 | 3417 | 23 | ABL17814 |
| 13 | 19 | 1.3 | 6006 | 23 | ABL17820 |
| 14 | 19 | 1.3 | 6199 | 22 | AA040631 |
| 15 | 19 | 1.3 | 6199 | 22 | AA040631 |
| 16 | 19 | 1.3 | 6199 | 22 | AA040631 |
| 17 | 18 | 1.2 | 34980 | 22 | AAH68532 |
| 18 | 18 | 1.2 | 365 | 22 | AAH57360 |
| 19 | 18 | 1.2 | 544 | 22 | AAK92455 |
| 20 | 18 | 1.2 | 843 | 23 | ABL21765 |
| 21 | 18 | 1.2 | 1125 | 21 | AAA75090 |
| 22 | 18 | 1.2 | 1146 | 22 | AAF61289 |
| 23 | 18 | 1.2 | 1262 | 21 | AAZ32926 |
| 24 | 18 | 1.2 | 1299 | 20 | AAK19492 |
| 25 | 18 | 1.2 | 1410 | 21 | AAK77073 |
| 26 | 18 | 1.2 | 1488 | 16 | AAQ86275 |
| 27 | 18 | 1.2 | 1532 | 21 | AAZ46950 |
| 28 | 18 | 1.2 | 1560 | 23 | AA588781 |
| 29 | 18 | 1.2 | 1560 | 23 | AA591130 |
| 30 | 18 | 1.2 | 1710 | 21 | AAF07491 |
| 31 | 18 | 1.2 | 1839 | 17 | AAH18897 |
| 32 | 18 | 1.2 | 1938 | 22 | AAH17691 |
| 33 | 18 | 1.2 | 1968 | 23 | ABL19437 |
| 34 | 18 | 1.2 | 1999 | 23 | ABL20087 |
| 35 | 18 | 1.2 | 2282 | 22 | AAK94158 |
| 36 | 18 | 1.2 | 2240 | 22 | AA531168 |
| 37 | 18 | 1.2 | 2625 | 22 | AA544597 |
| 38 | 18 | 1.2 | 2644 | 22 | AA544769 |
| 39 | 18 | 1.2 | 3192 | 23 | ABL21764 |
| 40 | 18 | 1.2 | 4113 | 23 | ABL20086 |
| 41 | 18 | 1.2 | 4633 | 23 | ABL05694 |
| 42 | 18 | 1.2 | 4727 | 23 | ABL19436 |
| 43 | 18 | 1.2 | 6412 | 24 | AA561146 |
| 44 | 18 | 1.2 | 9729 | 22 | AA514089 |
| 45 | 18 | 1.2 | 9826 | 22 | AA514085 |

ALIGNMENTS

RESULT 1

AA02692

ID AA02692 standard; DNA: 1464 BP.

XX AA02692;

DT 02-MAY-2001 (first entry)

DE Exophiala spinifera aldehyde dehydrogenase coding sequence.

XX Fumonisin; degradative enzyme; transport enzyme; flavin monooxygenase;
XX aldehyde dehydrogenase; permease; p-glycoprotein; fumonisin esterase;
KW detoxification; mycotoxin; animal feed; human feed; silage;
KW transgenic plant; transgenic animal; microbial spray; ds.

XX Exophiala spinifera.

XX Key Location/Qualifiers

FT CDS 1..1464

FT /tag= a

FT /product= "E. spinifera aldehyde dehydrogenase"

XX WO200105980-A1.

XX 25-JAN-2001.

XX 14-JUL-1999; 99WO-US15824.

XX 14-JUL-1999; 99WO-US15824.

XX (PION-) PIONEER HI-BRED INT INC.

XX (CURA-) CURAGEN CORP.

XX Duwick JP, Maddox J, Gilliam J, Folkerts O, Crasta OR;
 XX WPI; 2001-147345/15.
 DR P-PSDB; AAY72635.
 XX Novel polynucleotides encoding Exophiala degradative or transport
 PT enzyme which is useful for detoxifying fumonisin or structurally
 PT related mycotoxin during processing of grain for human or animal food
 XX consumption -
 PS Claim 1b; Page 67-68; 90pp; English.
 XX The patent discloses novel polynucleotides encoding Exophiala spinifera
 CC fumonisin degradative or transport enzymes such as flavin monooxygenase,
 CC aldehyde dehydrogenase, permease and p-glycoprotein. A fumonisin esterase
 CC enzyme with at least one fumonisin degradative enzyme is useful for
 CC detoxifying fumonisin or a structurally related mycotoxin present in
 CC harvested grain, on application to a plant or to harvested grain
 CC during processing, or to processed grain that is to be used as animal
 CC or human feed, or as a silage. Nucleotide fragments of the present
 CC invention are useful as probes and primers. They can be introduced
 CC into microorganisms that multiply on plants to deliver enzymes to
 CC potential target crops. The genes encoding the degrading enzymes are
 CC introduced via a vector into a microbial host and the transformed host
 CC is supplied to the environment, plants or animals for reducing the
 CC pathogenicity of a fungus producing fumonisin. The genes of the
 CC invention are fermented in a bacterial host and the resulting bacteria
 CC is processed and used as a microbial spray. The nucleotide sequences
 CC can be used alone or in combination to engineer microbes or other
 CC organisms to metabolize fumonisin and resist its toxic effects.
 CC The present DNA sequence is the aldehyde dehydrogenase coding sequence
 CC from Exophiala spinifera.
 XX
 SQ Sequence 1464 BP; 346 A; 402 C; 406 G; 310 T; 0 other;

Query Match 100.0%; Score 1464; DB 22; Length 1464;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 atggttttttgcgtgacgaatacaagagtgaactcttcatcaacaatgaattgcttc 60
 Db 1 atggttttttgcgtgacgaatacaagagtgaactcttcatcaacaatgaattgcttc 60
 QY 61 tccaaaggggtccgagagattacgctcacgacccggtggacgaatccaccgttgccact 120
 Db 61 tccaaaggggtccgagagattacgctcacgacccggtggacgaatccaccgttgccact 120
 QY 121 gatgttcacgtgcccacgcccagatgctgcacagtcagtagccggttcggtcagggc 180
 Db 121 gatgttcacgtgcccacgcccagatgctgcacagtcagtagccggttcggtcagggc 180
 QY 181 gtcaaaaaggcccatgggaagattcacaggtgacacacgcccggcggtgcattcaag 240
 Db 181 gtcaaaaaggcccatgggaagattcacaggtgacacacgcccggcggtgcattcaag 240
 QY 241 ttgcggacgtcccgagagaacgcccgaagaagtcgctgctggagtcgctgcccacc 300
 Db 241 ttgcggacgtcccgagagaacgcccgaagaagtcgctgctggagtcgctgcccacc 300
 QY 301 ggtagaccggtgtcgatgacactatttcgacattcccaaacatggtctccgtgttcgc 360
 Db 301 ggtagaccggtgtcgatgacactatttcgacattcccaaacatggtctccgtgttcgc 360
 QY 361 tactatgcagggtggcccgacagatcgcggaaaacaccttcccagggaacacggaag 420
 Db 361 tactatgcagggtggcccgacagatcgcggaaaacaccttcccagggaacacggaag 420
 QY 421 ccgaattggcgttaacagccgatgggggtgtgtgctggattccagctggaacgcgact 480
 Db 421 ccgaattggcgttaacagccgatgggggtgtgtgctggattccagctggaacgcgact 480

RESULT 2
 ABL10879
 ID ABL10879 standard; cDNA; 1980 BP.
 XX

QY 481 tttcttacgtcgttggaagatagccccccgctccgcccggctgctcttcatttc 540
 Db 481 tttcttacgtcgttggaagatagccccccgctccgcccggctgctcttcatttc 540
 QY 541 aaagcctcgagagaaatccccgcgtggcggttctggtgctcgtctctcttcgcagaagcc 600
 Db 541 aaagcctcgagagaaatccccgcgtggcggttctggtgctcgtctctcttcgcagaagcc 600
 QY 601 ggaattccctcctgagatcgctgcagttcctcactggaacgagtcgagggcgagcattg 660
 Db 601 ggaattccctcctgagatcgctgcagttcctcactggaacgagtcgagggcgagcattg 660
 QY 661 ggcgtcacatggacattgcgaagatcagcttcacagatctgctgcggtgagcgcgcgc 720
 Db 661 ggcgtcacatggacattgcgaagatcagcttcacagatctgctgcggtgagcgcgcgc 720
 QY 721 gtaagcaagcaacactcaagtccaacatgaagcgcgtcactctagaaactgggggaaaag 780
 Db 721 gtaagcaagcaacactcaagtccaacatgaagcgcgtcactctagaaactgggggaaaag 780
 QY 781 ccaacacatgcttctcaacgaagctcctcgaacgagtcgagggggaatcggaagat 840
 Db 781 ccaacacatgcttctcaacgaagctcctcgaacgagtcgagggggaatcggaagat 840
 QY 841 ttctcaaaattcgggcaaaatttggtgtcccccctcctctgttctgtagtgcagatgggaaat 900
 Db 841 ttctcaaaattcgggcaaaatttggtgtcccccctcctctgttctgtagtgcagatgggaaat 900
 QY 901 tttagcggagaattccatggagtcctcgaatggctcattttggaggtgctcagagatggcct 960
 Db 901 tttagcggagaattccatggagtcctcgaatggctcattttggaggtgctcagagatggcct 960
 QY 961 ggcgaagaacccattggaacccaagagagcagcatggtccctctcgtgcagaagtcaccagtac 1020
 Db 961 ggcgaagaacccattggaacccaagagagcagcatggtccctctcgtgcagaagtcaccagtac 1020
 QY 1021 gacagagcttgggttaacattgacgttggcaagatcacgcgcagctccctcactggcgctt 1080
 Db 1021 gacagagcttgggttaacattgacgttggcaagatcacgcgcagctccctcactggcgctt 1080
 QY 1081 ggtagaagaaggcgcaaaaggatcgcgattgaaacccagcagatatttgcatacccaacca 1140
 Db 1081 ggtagaagaaggcgcaaaaggatcgcgattgaaacccagcagatatttgcatacccaacca 1140
 QY 1141 ggcagcaaaattggttggagagatcttggcccgctcttccatttaagacgcttcaag 1200
 Db 1141 ggcagcaaaattggttggagagatcttggcccgctcttccatttaagacgcttcaag 1200
 QY 1201 acggaagaaggccattgagattgccaatgacacgacttatgggctagcctcggtcatt 1260
 Db 1201 acggaagaaggccattgagattgccaatgacacgacttatgggctagcctcggtcatt 1260
 QY 1261 tataccaatctctcaacaggggtctcgtgtcgttcggtcgcgcgcgcgcgcgcgcgcgc 1320
 Db 1261 tataccaatctctcaacaggggtctcgtgtcgttcggtcgcgcgcgcgcgcgcgcgcgc 1320
 QY 1321 tcgacataacttccccctttatccccgagacacaaactccggttggcggtgataaacaatcg 1380
 Db 1321 tcgacataacttccccctttatccccgagacacaaactccggttggcggtgataaacaatcg 1380
 QY 1381 ggcctcagggcagagactgagggcgaagaaggctcgaaggcgttacttggagcccaagaccatt 1440
 Db 1381 ggcctcagggcagagactgagggcgaagaaggctcgaaggcgttacttggagcccaagaccatt 1440
 QY 1441 aatatccacgtcaacatagatga 1464
 Db 1441 aatatccacgtcaacatagatga 1464

AC ABL10879;
 DT 26-MAR-2002 (first entry)
 XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 27119.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX Drosophila melanogaster.
 OS Drosophila melanogaster.
 PN WO200171042-A2.
 XX 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 PF 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2001; 2001WO-US09231.
 PR 11-JUL-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX Claim 1; SEQ ID NO 27119; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 1980 BP; 471 A; 527 C; 540 G; 442 T; 0 other;
 SQ
 Query Match 1.4%; Score 20; DB 23; Length 1980;
 Best Local Similarity 100.0%; Pred. No. 8.1;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1159 gagagatcttggcccccgt 1178
 Db 1681 gagagatcttggcccccgt 1700
 RESULT 3
 ABL24684/c
 ID ABL24684 standard; DNA; 3278 BP.
 XX AC ABL24684;
 XX 26-MAR-2002 (first entry)
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 25525.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX Drosophila melanogaster.
 OS Drosophila melanogaster.
 PN WO200171042-A2.

XX 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2001; 2001WO-US09231.
 PF 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2001; 2001WO-US09231.
 PR 11-JUL-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX Claim 1; SEQ ID NO 25525; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 3278 BP; 922 A; 729 C; 679 G; 948 T; 0 other;
 SQ
 Query Match 1.4%; Score 20; DB 23; Length 3278;
 Best Local Similarity 100.0%; Pred. No. 8.1;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1159 gagagatcttggcccccgt 1178
 Db 3046 GAGGAGATCTTTGGCCCCGT 3027
 RESULT 4
 ABL10878/c
 ID ABL10878 standard; cDNA; 6153 BP.
 XX AC ABL10878;
 XX 26-MAR-2002 (first entry)
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 27116.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX Drosophila melanogaster.
 OS Drosophila melanogaster.
 PN WO200171042-A2.
 XX 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 PF 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2001; 2001WO-US09231.
 PR 11-JUL-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 XX

DR P-PSDB; ABB66775.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 XX Claim 1; SEQ ID NO 27116; 21pp + Sequence Listing; English.
 XX
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 6153 BP; 1689 A; 1322 C; 1355 G; 1787 T; 0 other;

Query Match 1.4%; Score 20; DB 23; Length 6153;
 Best Local Similarity 100.0%; Pred. No. 8;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1159 gagagatcttggcccggt 1178
 |||||
 Db 1300 GAGGAGATCTTTGGCCCCGT 1281

RESULT 5
 AAC03035
 ID AAC03035 standard; cDNA; 286 BP.
 XX
 AC AAC03035;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein 5' EST, SEQ ID NO: 3033.
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1033401-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 21-FEB-2000; 2000EP-0200610.
 XX
 PR 26-FEB-1999; 99US-0122487.
 XX
 PA (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;
 XX WPI; 2000-500381/45.
 DR P-PSDB; AAG03029.
 XX
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 XX Claim 1; SEQ ID 3033; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or poly(A)⁺ RNAs
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are

CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.
 XX
 XX Sequence 286 BP; 59 A; 89 C; 50 G; 84 T; 4 other;

Query Match 1.3%; Score 19; DB 21; Length 286;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 528 ctcttcattcttcaggcc 546
 |||||
 Db 116 ctcttcattcttcaggcc 134

RESULT 6
 AAF68055
 ID AAF68055 standard; DNA; 578 BP.
 XX
 AC AAF68055;
 XX
 DT 11-APR-2001 (first entry)
 XX
 DE Corynebacterium glutamicum MCT protein encoding DNA SEQ ID NO:625.
 XX
 KW Corynebacterium glutamicum; Brevibacterium lactofermentum; MCT;
 KW membrane construction and membrane transport protein; petroleum spill;
 KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
 KW identification; microorganism; fine chemical production; transformation;
 KW genome mapping; genetic engineering; ds.

XX Corynebacterium glutamicum.

XX WO200100805-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-IB00926.

XX 25-JUN-1999; 99US-0141031.

PR 08-JUL-1999; 99DE-1031454.

PR 08-JUL-1999; 99DE-1031478.

PR 08-JUL-1999; 99DE-1031563.

PR 09-JUL-1999; 99DE-1032122.

PR 09-JUL-1999; 99DE-1032124.

PR 09-JUL-1999; 99DE-1032125.

PR 09-JUL-1999; 99DE-1032128.

PR 09-JUL-1999; 99DE-1032180.

PR 09-JUL-1999; 99DE-1032182.

PR 09-JUL-1999; 99DE-1032190.

PR 09-JUL-1999; 99DE-1032191.

PR 09-JUL-1999; 99DE-1032209.

PR 09-JUL-1999; 99DE-1032212.

PR 09-JUL-1999; 99DE-1032227.

PR 09-JUL-1999; 99DE-1032228.

PR 09-JUL-1999; 99DE-1032229.

PR 09-JUL-1999; 99DE-1032230.

PR 14-JUL-1999; 99DE-1032927.

PR 14-JUL-1999; 99DE-1033005.

PR 14-JUL-1999; 99DE-1033006.

PR 27-AUG-1999; 99DE-1040764.

PR 27-AUG-1999; 99DE-1040765.

PR 27-AUG-1999; 99DE-1040766.

PR 27-AUG-1999; 99DE-1040830.

PR 27-AUG-1999; 99DE-1040831.

PR 27-AUG-1999; 99DE-1040832.

PR 27-AUG-1999; 99DE-1040833.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041395.
 PR 03-SEP-1999; 99DE-1042077.
 PR 03-SEP-1999; 99DE-1042078.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042088.
 XX (BADI) BASF AG.
 XX
 PI Pompejus M, Kroeger B, Schroeder H, zelder O, Haberhauer G;
 XX
 DR WPI: 2001-071486/08.
 DR P-PSDB; AAB76822.
 XX
 XX Corynebacterium glutamicum nucleic acids encoding membrane construction
 PT and membrane transport proteins or their portions, useful for typing or
 PT identifying C. glutamicum or related bacteria, and as markers for
 PT transformation.
 XX
 PS Claim 3; Page 1048-1049; 1119pp; English.
 XX
 CC AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
 CC construction and membrane transport (MCT) proteins given in AAB76510 to
 CC AAB76847. The MCT nucleic acids and proteins are useful in the
 CC identification of microorganisms which can be used to produce fine
 CC chemicals, for modulating fine chemical production in C. glutamicum or
 CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
 CC identification of C. glutamicum or related bacteria, as reference points
 CC for mapping C. glutamicum genome, and as markers for transformation.
 CC AAF68082 and AAF68082 represent sequencing primers which are used in an
 CC example from the present invention.
 XX
 SQ Sequence 578 BP; 146 A; 182 C; 136 G; 114 T; 0 other;

 Query Match 1.3%; Score 19; DB 22; Length 578;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 1215 cattgagattgccaatgac 1233
 |||||
 Db 456 cattgagattgccaatgac 474

 RESULT 7
 AAF22587/c
 ID AAF22587 standard; cDNA; 600 BP.
 XX
 AC AAF22587;
 XX
 DT 26-MAR-2001 (first entry)
 XX
 DE Human breast cancer associated antigen nucleotide sequence SEQ ID NO.166.
 XX
 KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
 KW cancer associated antigen; cytostatic; cancer vaccine; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200073801-A2.
 XX
 PD 07-DEC-2000.
 XX
 PF 26-MAY-2000; 2000WO-US14749.
 XX
 PR 28-MAY-1999; 99US-0136526.
 PR 10-SEP-1999; 99US-0153454.
 XX
 PA (LUDW-) LUDWIG INST CANCER RES.
 XX
 PI Obata Y;

XX WPI: 2001-025274/03.
 DR
 XX Nucleic acids encoding breast, gastric and prostate cancer associated
 PT antigen precursors, useful for diagnosing and treating a condition
 PT characterized by expression of an abnormal amount of a protein, e.g.
 PT cancer -
 XX
 PS Claim 50; Page 328; 799pp; English.
 XX
 CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
 CC represent nucleotide sequences encoding human breast, gastric and
 CC prostate cancer associated antigen precursors (CAAP) respectively.
 CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
 CC represent human breast, gastric and prostate CAAP protein sequence
 CC respectively. CAAPs have cytostatic activity and can be used in the
 CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
 CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
 CC condition characterised by expression of an abnormal amount of a protein,
 CC e.g. cancer.
 XX
 SQ Sequence 600 BP; 103 A; 116 C; 113 G; 193 T; 75 other;

 Query Match 1.3%; Score 19; DB 22; Length 600;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 1325 tcaactcccccctttatccc 1343
 |||||
 Db 451 TCAACTTCCCTTATCC 433

 RESULT 8
 ABL17815
 ID ABL17815 standard; DNA; 828 BP.
 XX
 AC ABL17815;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 4918.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 4918; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 828 BP; 161 A; 241 C; 237 G; 189 T; 0 other;

Query Match 1.3%; Score 19; DB 23; Length 828;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 gtgcaggcggtcaaaaagg 190
 |||||

Db 778 gtgcaggcggtcaaaaagg 796
 |||||

RESULT 9
 AAF22491/C
 ID AAF22491 standard; cDNA; 1015 BP.

XX AC AAF22491;

XX DT 26-MAR-2001 (first entry)

XX DE Human breast cancer associated antigen nucleotide sequence SEQ ID NO:70.

XX XX Human; breast cancer; gastric cancer; prostate cancer; diagnosis;

KW cancer associated antigen; cytostatic; cancer vaccine; ss.

XX OS Homo sapiens.

XX PN WO200073801-A2.

XX PD 07-DEC-2000.

XX PF 26-MAY-2000; 2000WO-US14749.

XX PR 28-MAY-1999; 99US-0136526.

XX PR 10-SEP-1999; 99US-0153454.

XX PA (LUDW-) LUDWIG INST CANCER RES.

XX PI Obata Y;

XX PS WPI; 2001-025274/03.

XX PT Nucleic acids encoding breast, gastric and prostate cancer associated
 PT antigen precursors, useful for diagnosing and treating a condition
 PT characterized by expression of an abnormal amount of a protein, e.g.
 PT cancer -

XX Claim 50; Page 297; 799pp; English.

XX PS AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
 CC represent nucleotide sequences encoding human breast, gastric and
 CC prostate cancer associated antigen precursors (CAAP), respectively.
 CC AAF63232 to AAF63467, AAF63468 to AAF63721 and AAF63722 to AAF63970
 CC represent human breast, gastric and prostate CAAP protein sequence
 CC respectively. CAAPs have cytostatic activity and can be used in the
 CC reduction of cancer vaccines. The human CAAP proteins, peptides, nucleic
 CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
 CC condition characterised by expression of an abnormal amount of a protein,
 CC e.g. cancer.

XX SQ Sequence 1015 BP; 192 A; 179 C; 186 G; 340 T; 118 other;

Query Match 1.3%; Score 19; DB 22; Length 1015;
 Best Local Similarity 100.0%; Pred. No. 26;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1325 tcaacttccctttatccc 1343
 |||||

Db 451 TCAACTTCCCTTTATCCC 433
 |||||

RESULT 10
 AAH67550
 ID AAH67550 standard; DNA; 1887 BP.

XX AC AAH67550;

XX DT 26-SEP-2001 (first entry)

XX DE C glutamicum coding sequence fragment SEQ ID NO: 2585.

KW Corynebacterium bacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis; ds.

XX OS Corynebacterium glutamicum.

XX PN EP1108790-A2.

XX PD 20-JUN-2001.

XX PF 18-DEC-2000; 2000EP-0127688.

XX PR 16-DEC-1999; 99JP-0377484.

XX PR 07-APR-2000; 2000JP-0159162.

XX PR 03-AUG-2000; 2000JP-0280988.

XX PA (KYOW) KYOWA HAKKO KOGYO KK.

XX PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX DR WPI; 2001-376931/40.

XX PS P-FSDB; AAG92331.

XX PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene

XX PS Claim 8; SEQ ID NO: 2585; 246pp + Sequence Listing; English.

XX CC The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium glutamicum glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of corynebacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Corynebacterium, and identifying a homologue of a gene derived
 CC from corynebacterium. Corynebacterium bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.

XX CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

XX SQ Sequence 1887 BP; 338 A; 576 C; 498 G; 475 T; 0 other;

Query Match 1.3%; Score 19; DB 22; Length 1887;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1215 cattgagattgccaatgac 1233
 |||||

Db 1788 cattgagattgccaatgac 1806
 |||||

Query Match 1.3%; Score 19; DB 23; Length 3417;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 172 gtgcaggcggtcaaaaagg 190
|||||

Db 2367 gtgcaggcggtcaaaaagg 2385
|||||

RESULT 13

ABL17820
ID ABL17820 standard; DNA; 6006 BP.

XX AC ABL17820;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 4933.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical; gene; ds.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions -

PS Claim 1; SEQ ID NO 4933; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
sequences (ABL01840-ABL16175) and the encoded proteins
(AB5737-AB572072).

CC The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pot_sequences.

XX SQ Sequence 6006 BP; 1478 A; 1504 C; 1511 G; 1513 T; 0 other;

Query Match 1.3%; Score 19; DB 23; Length 6006;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 172 gtgcaggcggtcaaaaagg 190
|||||

Db 3724 gtgcaggcggtcaaaaagg 3742
|||||

RESULT 14

AAS40631

ID AAS40631 standard; DNA; 6199 BP.

XX AC AAS40631;

XX DT 17-DEC-2001 (first entry)

XX DE DNA encoding human prostate cancer antigen, Seq ID NO 783.

XX KW Human; prostate cancer antigen; cytostatic; uropathic; diagnostic;
reproductive system; chromosomal marker; forensic; urinary disorder;
chronic nephritis; blood-related disorder; thrombosis; ds.

XX OS Homo sapiens.

XX PN WO200155316-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01328.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 24-FEB-2000; 2000US-0184664.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

XX PR 17-MAR-2000; 2000US-0190076.

XX PR 18-APR-2000; 2000US-0198123.

XX PR 19-MAY-2000; 2000US-0205515.

XX PR 07-JUN-2000; 2000US-0209467.

XX PR 28-JUN-2000; 2000US-0214886.

XX PR 30-JUN-2000; 2000US-0215135.

XX PR 07-JUL-2000; 2000US-0216647.

XX PR 11-JUL-2000; 2000US-0216880.

XX PR 11-JUL-2000; 2000US-0217487.

XX PR 11-JUL-2000; 2000US-0217496.

XX PR 14-JUL-2000; 2000US-0218290.

XX PR 28-JUL-2000; 2000US-0220963.

XX PR 26-JUL-2000; 2000US-0220964.

XX PR 14-AUG-2000; 2000US-0224518.

XX PR 14-AUG-2000; 2000US-0224519.

XX PR 14-AUG-2000; 2000US-0225213.

XX PR 14-AUG-2000; 2000US-0225214.

XX PR 14-AUG-2000; 2000US-0225266.

XX PR 14-AUG-2000; 2000US-0225267.

XX PR 14-AUG-2000; 2000US-0225268.

XX PR 14-AUG-2000; 2000US-0225270.

XX PR 14-AUG-2000; 2000US-0225447.

XX PR 14-AUG-2000; 2000US-0225757.

XX PR 14-AUG-2000; 2000US-0225758.

XX PR 14-AUG-2000; 2000US-0225759.

XX PR 18-AUG-2000; 2000US-0226279.

XX PR 22-AUG-2000; 2000US-0226681.

XX PR 22-AUG-2000; 2000US-0226868.

XX PR 22-AUG-2000; 2000US-0227182.

XX PR 23-AUG-2000; 2000US-0227009.

XX PR 30-AUG-2000; 2000US-0228924.

XX PR 01-SEP-2000; 2000US-0229287.

XX PR 01-SEP-2000; 2000US-0229343.

XX PR 01-SEP-2000; 2000US-0229344.

XX PR 01-SEP-2000; 2000US-0229345.

XX PR 05-SEP-2000; 2000US-0229509.

XX PR 05-SEP-2000; 2000US-0229513.

XX PR 06-SEP-2000; 2000US-0230437.

XX PR 06-SEP-2000; 2000US-0230438.

XX PR 08-SEP-2000; 2000US-0231242.

XX PR 08-SEP-2000; 2000US-0231243.

XX PR 08-SEP-2000; 2000US-0231244.

XX PR 08-SEP-2000; 2000US-0231413.

XX PR 08-SEP-2000; 2000US-0231414.

XX PR 08-SEP-2000; 2000US-0232080.

XX PR 08-SEP-2000; 2000US-0232081.

XX PR 12-SEP-2000; 2000US-0231968.

XX PR 14-SEP-2000; 2000US-0232397.

| | | |
|----|--------------|------------------|
| PR | 24-FEB-2000; | 2000US-01846660; |
| PR | 02-MAR-2000; | 2000US-01863500; |
| PR | 16-NAR-2000; | 2000US-01898740; |
| PR | 17-NAR-2000; | 2000US-01900760; |
| PR | 18-APR-2000; | 2000US-01981213; |
| PR | 19-MAY-2000; | 2000US-02055115; |
| PR | 19-JUN-2000; | 2000US-02094670; |
| PR | 28-JUN-2000; | 2000US-02148860; |
| PR | 30-JUN-2000; | 2000US-02151395; |
| PR | 07-JUL-2000; | 2000US-02166647; |
| PR | 07-JUL-2000; | 2000US-02168800; |
| PR | 11-JUL-2000; | 2000US-02174880; |
| PR | 11-JUL-2000; | 2000US-02174960; |
| PR | 14-JUL-2000; | 2000US-02182990; |
| PR | 26-JUL-2000; | 2000US-02209653; |
| PR | 26-JUL-2000; | 2000US-02209664; |
| PR | 14-AUG-2000; | 2000US-02245318; |
| PR | 14-AUG-2000; | 2000US-02245319; |
| PR | 14-AUG-2000; | 2000US-02252113; |
| PR | 14-AUG-2000; | 2000US-02252114; |
| PR | 14-AUG-2000; | 2000US-02252660; |
| PR | 14-AUG-2000; | 2000US-02252670; |
| PR | 14-AUG-2000; | 2000US-02252680; |
| PR | 14-AUG-2000; | 2000US-02252770; |
| PR | 14-AUG-2000; | 2000US-02254547; |
| PR | 14-AUG-2000; | 2000US-02257557; |
| PR | 14-AUG-2000; | 2000US-02257578; |
| PR | 14-AUG-2000; | 2000US-02257590; |
| PR | 18-AUG-2000; | 2000US-02262790; |
| PR | 22-AUG-2000; | 2000US-02266810; |
| PR | 22-AUG-2000; | 2000US-02268680; |
| PR | 23-AUG-2000; | 2000US-02271820; |
| PR | 23-AUG-2000; | 2000US-02270009; |
| PR | 30-AUG-2000; | 2000US-02289240; |
| PR | 01-SEP-2000; | 2000US-02292870; |
| PR | 01-SEP-2000; | 2000US-02293430; |
| PR | 01-SEP-2000; | 2000US-02293440; |
| PR | 01-SEP-2000; | 2000US-02293445; |
| PR | 05-SEP-2000; | 2000US-02295909; |
| PR | 05-SEP-2000; | 2000US-02295913; |
| PR | 06-SEP-2000; | 2000US-02304337; |
| PR | 06-SEP-2000; | 2000US-02304380; |
| PR | 08-SEP-2000; | 2000US-02311420; |
| PR | 08-SEP-2000; | 2000US-02311423; |
| PR | 08-SEP-2000; | 2000US-02312440; |
| PR | 08-SEP-2000; | 2000US-02314113; |
| PR | 08-SEP-2000; | 2000US-02314140; |
| PR | 08-SEP-2000; | 2000US-02320800; |
| PR | 08-SEP-2000; | 2000US-02320810; |
| PR | 12-SEP-2000; | 2000US-02319680; |
| PR | 14-SEP-2000; | 2000US-02323970; |
| PR | 14-SEP-2000; | 2000US-02323998; |
| PR | 14-SEP-2000; | 2000US-02323999; |
| PR | 14-SEP-2000; | 2000US-02324000; |
| PR | 14-SEP-2000; | 2000US-02324010; |
| PR | 14-SEP-2000; | 2000US-02330663; |
| PR | 14-SEP-2000; | 2000US-02330664; |
| PR | 14-SEP-2000; | 2000US-02330665; |
| PR | 21-SEP-2000; | 2000US-02342223; |
| PR | 21-SEP-2000; | 2000US-02342774; |
| PR | 21-SEP-2000; | 2000US-02343997; |
| PR | 25-SEP-2000; | 2000US-02344998; |
| PR | 26-SEP-2000; | 2000US-02354884; |
| PR | 27-SEP-2000; | 2000US-02358934; |
| PR | 27-SEP-2000; | 2000US-02358936; |
| PR | 29-SEP-2000; | 2000US-02363277; |
| PR | 29-SEP-2000; | 2000US-02363287; |
| PR | 29-SEP-2000; | 2000US-02363668; |
| PR | 29-SEP-2000; | 2000US-02363669; |
| PR | 29-SEP-2000; | 2000US-02363700; |
| PR | 02-OCT-2000; | 2000US-02368002; |
| PR | 02-OCT-2000; | 2000US-02370337; |
| PR | 02-OCT-2000; | 2000US-02703038; |

CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.

XX
SQ Sequence 6199 BP; 1516 A; 1686 C; 1432 G; 1565 T; 0 other;

Query Match 1.3%; Score 19; DB 22; Length 6199;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1211 aggccattgagattgccaa 1229
|||||
DB 1648 aggccattgagattgccaa 1666

Search completed: April 27, 2002, 05:28:23
Job time: 13786 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2002, 05:23:55 ; Search time 7820.58 Seconds
(without alignments)
3917.414 Million cell updates/sec

Title: US-09-882-694A-4
Perfect score: 1464
Sequence: 1 atggtttcttcgcctgacga.....tccagtcacacatagatga 1464

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0
Searched: 1797656 seqs, 10463268293 residues
Word size : 0
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_btg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|-------|--------|----|----|-------------|
|------------|-------|-------|-------|--------|----|----|-------------|

| | | | | | | | | | | | |
|----|------|-------|--------|----|--------------|--------------|-----------------------------------|---------|-----|--------|-----------------|
| 1 | 1464 | 100.0 | 1464 | 6 | AX076845 | AX076845 | Sequence 4 from Patent WO0105980. | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| 2 | 24 | 1.6 | 4158 | 8 | ABY17825 | ABY17825 | | | | | |
| 3 | 22 | 1.5 | 157201 | 8 | AP003048 | AP003048 | | | | | |
| 4 | 22 | 1.5 | 260050 | 1 | AL603643 | AL603643 | | | | | |
| 5 | 21 | 1.4 | 3897 | 9 | HSU62769 | HSU62769 | | | | | |
| 6 | 21 | 1.4 | 4226 | 9 | HSU131024 | HSU131024 | | | | | |
| 7 | 21 | 1.4 | 10498 | 9 | AB008831 | AB008831 | | | | | |
| 8 | 21 | 1.4 | 102910 | 9 | AL356308 | AL356308 | | | | | |
| 9 | 21 | 1.4 | 131805 | 9 | AC008850 | AC008850 | | | | | |
| 10 | 21 | 1.4 | 138904 | 2 | AF003928 | AF003928 | | | | | |
| 11 | 21 | 1.4 | 147344 | 2 | AC008907 | AC008907 | | | | | |
| 12 | 21 | 1.4 | 191415 | 2 | AL627302 | AL627302 | | | | | |
| 13 | 21 | 1.4 | 251664 | 2 | AC008908 | AC008908 | | | | | |
| 14 | 20 | 1.4 | 481 | 14 | FVNS5GAI | FVNS5GAI | | | | | |
| 15 | 20 | 1.4 | 2001 | 1 | AB009654 | AB009654 | | | | | |
| 16 | 20 | 1.4 | 11553 | 1 | AE005775 | AE005775 | | | | | |
| 17 | 20 | 1.4 | 20701 | 1 | AE006958 | AE006958 | | | | | |
| 18 | 20 | 1.4 | 29550 | 1 | MTV039 | MTV039 | | | | | |
| 19 | 20 | 1.4 | 31087 | 2 | AC014839 | AC014839 | | | | | |
| 20 | 20 | 1.4 | 48986 | 2 | AC105976 | AC105976 | | | | | |
| 21 | 20 | 1.4 | 58005 | 2 | AC109128 | AC109128 | | | | | |
| 22 | 20 | 1.4 | 73635 | 2 | AC100991 | AC100991 | | | | | |
| 23 | 20 | 1.4 | 79722 | 2 | AC018384 | AC018384 | | | | | |
| 24 | 20 | 1.4 | 110000 | 2 | LMFICHR25_02 | LMFICHR25_02 | | | | | |
| 25 | 20 | 1.4 | 126059 | 8 | AP003197 | AP003197 | | | | | |
| 26 | 20 | 1.4 | 129007 | 2 | AC106094 | AC106094 | | | | | |
| 27 | 20 | 1.4 | 138025 | 9 | HSUJ20N4 | HSUJ20N4 | | | | | |
| 28 | 20 | 1.4 | 162501 | 2 | AC097791 | AC097791 | | | | | |
| 29 | 20 | 1.4 | 166048 | 9 | AC013725 | AC013725 | | | | | |
| 30 | 20 | 1.4 | 167449 | 2 | AC025073 | AC025073 | | | | | |
| 31 | 20 | 1.4 | 172538 | 2 | AC041016 | AC041016 | | | | | |
| 32 | 20 | 1.4 | 173333 | 2 | AC025282 | AC025282 | | | | | |
| 33 | 20 | 1.4 | 174894 | 3 | AC007984 | AC007984 | | | | | |
| 34 | 20 | 1.4 | 175347 | 2 | AC009099 | AC009099 | | | | | |
| 35 | 20 | 1.4 | 177241 | 2 | AC102960 | AC102960 | | | | | |
| 36 | 20 | 1.4 | 187388 | 2 | AC094898 | AC094898 | | | | | |
| 37 | 20 | 1.4 | 188216 | 2 | AC092711 | AC092711 | | | | | |
| 38 | 20 | 1.4 | 225092 | 3 | AE003758 | AE003758 | | | | | |
| 39 | 20 | 1.4 | 251700 | 1 | AF000062 | AF000062 | | | | | |
| 40 | 19 | 1.3 | 578 | 6 | AX067043 | AX067043 | | | | | |
| 41 | 19 | 1.3 | 600 | 6 | AX053400 | AX053400 | | | | | |
| 42 | 19 | 1.3 | 1015 | 6 | AX053304 | AX053304 | | | | | |
| 43 | 19 | 1.3 | 1251 | 10 | AF233226 | AF233226 | | | | | |
| 44 | 19 | 1.3 | 1887 | 6 | AX122669 | AX122669 | | | | | |
| 45 | 19 | 1.3 | 2010 | 6 | AX067041 | AX067041 | | | | | |

ALIGNMENTS

| | | | | | | |
|------------|------------|-----------------------------------|---------|-----|--------|-----------------|
| RESULT 1 | AX076845 | Sequence 4 from Patent WO0105980. | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| LOCUS | AX076845 | Sequence 4 from Patent WO0105980. | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| DEFINITION | AX076845 | Sequence 4 from Patent WO0105980. | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| ACCESSION | AX076845 | Sequence 4 from Patent WO0105980. | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| VERSION | AX076845.1 | GI:13121519 | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| KEYWORDS | AX076845.1 | GI:13121519 | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| SOURCE | AX076845.1 | GI:13121519 | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| ORGANISM | AX076845.1 | GI:13121519 | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| REFERENCE | AX076845.1 | GI:13121519 | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| AUTHORS | AX076845.1 | GI:13121519 | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| TITLE | AX076845.1 | GI:13121519 | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| JOURNAL | AX076845.1 | GI:13121519 | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| FEATURES | AX076845.1 | GI:13121519 | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| source | AX076845.1 | GI:13121519 | 1464 bp | DNA | linear | PAT 22-FEB-2001 |
| BASE COUNT | AX076845.1 | GI:13121519 | 1464 bp | DNA | linear | PAT 22-FEB-2001 |

ORIGIN

Query Match 100.0%; Score 1464; DB 6; Length 1464;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgggtcttctgctgacgaataacagagtgaaactcttctcatcaaatgaattcgtctcc 60
DB 1 ATGGTCTTCTCGCCTCACCATAACAGAGTGAACCTCTCATCAACAATGAATTCGTC 60

QY 61 tcaaggggtccagagattaaacgtctcagacaccgtggaggaatccacccttgccact 120
DB 61 TCAAGGGTCCGAGAGATTACGCTCACGAACCCGTGGGAGAAATCCACCCTGCCACT 120

QY 121 gatgttcacgtggccaacgcggccgatgctgcacagtgacagtcgctcgtgagggcg 180
DB 121 GATGTTTACGTTGGCCAAACGCCGCGATGTCGACAGTGCAGTGCCTTCGCTGCAGGCG 180

QY 181 gtcaaaaggccctatgaagattcacaggtgcacacgcgcgcgcgcgcgcgcgcgcgc 240
DB 181 GTCAAAAGGCCCTATGGAAGAGATTTCACAGGTGCACAAACGCCGCGCGTGCATGCTTAAG 240

QY 241 ttcgcggacctgcgcgagagaacgcgcgagaagctcgcctcgtcgtgagtcgctgcccacc 300
DB 241 TTCGCGGACCTGCCGAGAGAAGCCGCGAGAAGCTCGCTCGTCTGGAGTCGCTGCCACC 300

QY 301 gctagaccggtgcagatgatactcatttcgacattccaaacatgctcgcgtgttcgcg 360
DB 301 GGTAGACCCGCTGTCGATGATCACTCATTTTCGACATTCCTAAACATGCTCTCCGTTTCGC 360

QY 361 tactatgcagggtggccgacagatcgccggaagaccttcccgagagacaacgcgcaag 420
DB 361 TACTATGCAGGCTGGCGCCGACAAAGATCGCCGGAAGACCTTTCCCGAGGACAACGGCAAG 420

QY 421 ccgaattgctcagaccgagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
DB 421 CGGAATTGCGGTACGACCGGATGGGGGTGTGCTGTATTGCGCAGCTTGGAAACGCGACT 480

QY 481 ttctttacgtcgggtggaagatagccccgcctcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
DB 481 TTTCTTTACGTCGGTGGGAAGATAGCCCGCCCTCGCGCGCGGCTGCTCTTCATCTTC 540

QY 541 aaagctcgagaaatccccctggcgttctggccctcgcctcctctctctctctctctctct 600
DB 541 AAAGCTCGAGAAATCCCCGCTGGCGCTTCGCGCCCTCGCTCTCTCTCTCTCTCTCTCT 600

QY 601 ggattccctcctggagtcgtcagttcctcactggagcacgagtgacgggtgaaacattg 660
DB 601 GGATTCCCTCTCTGGAGTCGTGAGTTCCTCACTGGAGCACGAGTACGCGGTGAAGCAATTG 660

QY 661 gcgtcgacatggacattgcgaagatcagcttcacaagatctgtcgcggtgcccgcgc 720
DB 661 GCGTCGCACATGACATTCGGAAGATCAGCTTCACAAGATCTGTCCGGCGGTGGCCGCC 720

QY 721 gtaagcagaacactcaagtcacaactgaagcgcgtcactctagaactgggggaaag 780
DB 721 GTCAAGCAAGCAACACTCAAGTCCAACATGAAGCGCGTCACTCTAGAACTGGGGAAAG 780

QY 781 ccaacatcgtcttcaacgaagctcctcgaacgcgagtcggggaatcggaagat 840
DB 781 CCANACATCGCTTCAACGAAGCTCCTCTCGAACGGCAGTGGGGGAATCGGCAAGGAT 840

QY 841 ttctcaaatgggcaaatattgggtcccccctcctcttctgtagtcaatggggaat 900
DB 841 TTCTCAAAATTCGGGCAAAATTTGGTCCGCCCTCTCTGCTAGTCAATGGGGAAT 900

QY 901 ttacgaggaattccatggagtcgctcatgctcattgaggtgtcagatggctt 960
DB 901 TTAGCGGAGAAATTCATGGAGTCCGTCTGCTCATTTGGAGGCTGTCAAGATGGCTT 960

QY 961 ggccagaacccattggaacccaaaggagcgatggtcctctcgtcgacaagtcctcagtac 1020
DB 961 GGCCAGAACCCATTGGAACCCAAAGGAGCGCATGGTCCCTCTCGTCACAAGTCCCAAGTAC 1020

Db 961 GGCCAGAACCCATTGGAACCCAAAGAGCGCATGGTCCCTCTCGTCACAAGTCCCAAGTAC 1020

QY 1021 gcaagagtccttgggttaacaattgacgtttggcaaggatataccgcgcagctctcactcgtt 1080
Db 1021 GACAGAGTCTTGGGTAAACATTGACGTTGGCAAGGATACGGCGCAGCTCCTCACTGGCGTT 1080

QY 1081 ggtagaaggggcgacaaggattcgcgaltgaaccgaagatatttgcataatcccaaacca 1140
Db 1081 GGTAGAAAGGGCGACAAGGATTTCGCGATTGAACGACGATATTTGTCAATCCCAACCA 1140

QY 1141 ggcagcaaaattgtttgaggagatcttggcccgctcttgcattaaagcgttcaag 1200
Db 1141 GGCAGCAAAATTTGTTTGAAGAGATCTTTGGCCCGCTTGTCTCCATTAAAGACGTTCAAG 1200

QY 1201 acggaagaagggccattgagattgccaatgacacgacttaaggctagcctcgttcatt 1260
Db 1201 ACGGAAGAAGAGGCCATTGAGATTGCCAATGACACGACTTATGGGCTAGCCTCGTCAAT 1260

QY 1261 tatacaaatctctcaacaggggtctcgcgtctcgtcgcgcgcgcgcgcgcgcgcgcgcgc 1320
Db 1261 TATACCAATCTCTCAACAGGGGTCTCGGTGTCTCTCGTGGCGCTCGAGACCGTGGCGTC 1320

QY 1321 tcatcaactctcccttctcccccagacacaaactccggtttggcgcatgaaacaatcg 1380
Db 1321 TCGATCAACTTCCCTTTATCCCGGAGACAAACTCCGTTTGGCGCATGAACAATCG 1380

QY 1381 ggcctagggcagagactagggcgaagaaggctcgaaggcgtacttggagcccaagaccatt 1440
Db 1381 GGCTCAGGACAGAGCTAGGGCAAGAGGGTCAAGGGCTACTTTGGAGCCCAAGACCAATT 1440

QY 1441 aatatccacgtcaacatagagtga 1464
Db 1441 AATATCCACGTCAACATAGAGTGA 1464

RESULT 2

ABY17825 ABY17825 Agaricus bisporus alda 4158 bp DNA linear PLN 24-JUL-1998

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1. .4158

/Organism="Agaricus bisporus"

/strain="Horst H39"

/db_xref="taxon:5341"

Join(<532. .650,716. .729,787. .1170,1234. .1439,1494. .1657, 1713. .1881,1934. .2123,2190. .2302,2361. .2455,2516. .>2564),

/gene="alda"

532.2564

/gene="alda"

<532. .650

/gene="alda"

/number=1

Join(<532. .650,716. .729,787. .1170,1234. .1439,1494. .1657, 1713. .1881,1934. .2123,2190. .2302,2361. .2455,2516. .>2564),

CDS

```
/gene="aldA"
/EC_number=1.2.1.3"
/codon_start=1
/product="putative aldehyde dehydrogenase (NAD+)"
/protein_id="CAA76875.1"
/db_xref="GI:335324"
/db_xref="SPTREMBL:O74187"
/translation="MPSIFTHQMDTPVYKGSTSINTGLFINGEFVGVKNTTIDVVNP
ANGKITISATEADIDIAVEAAHKAFTTWGLNCSGSKRGDMLYKLAQMEKNIDQ
LSAIFALDNKTEFLWAKSVDSLISITKHYAGADKNFCQVETDEKXLTYSRHEPI
GVVGGIIPNPEPLMLANKIGPALATNCIVLKPSETPLSALRMCCALLOEACFPFGV
VNVYTGISTTGQALSSHKMKIDKVAFTGSTLVGRKVAEAAKSNLKNVTELGKSPV
VIFDDADEQSVNWTAGLEWNGHGOACCACTRIFFVQGGIYDRFLQKFTDKIKEIKLGD
PGLIGDGPQVSQIYDRIMSYTESGRAGATVHVGGRHGNEGFIQPTIFTDTTP
DKLIVKEIFPGVGAIVFKDQKEVIFQANDSNYGLAAVFSQDINKA1ETABAFKAG
TAWVNCANTIDAGVPFGYKQSGTGRELGELYALHNTNVKAVHVNLMKRN"
651..715
intron
/gene="aldA"
/number=1
/number=1
716..729
exon
/gene="aldA"
/number=2
730..786
intron
/gene="aldA"
/number=2
787..1170
exon
/gene="aldA"
/number=3
1171..1233
intron
/gene="aldA"
/number=3
1234..1439
exon
/gene="aldA"
/number=4
1440..1493
intron
/gene="aldA"
/number=4
1494..1657
exon
/gene="aldA"
/number=5
1658..1712
intron
/gene="aldA"
/number=5
1713..1881
exon
/gene="aldA"
/number=6
1882..1933
intron
/gene="aldA"
/number=6
1934..2123
exon
/gene="aldA"
/number=7
2124..2189
intron
/gene="aldA"
/number=7
2190..2302
exon
/gene="aldA"
/number=8
2303..2360
intron
/gene="aldA"
/number=8
2361..2455
exon
/gene="aldA"
/number=9
2456..2515
intron
/gene="aldA"
/number=9
2516..>2564
exon
/gene="aldA"
/number=10
<3094..3140
exon
/gene="echA"
/number=1
join(<3094..3140,3200..3435,3511..3840,3899..>3979)
/gene="echA"
3094..3979
/gene="echA"
join(<3094..3140,3200..3435,3511..3840,3899..3979)
/gene="echA"
/codon_start=2
/product="putative peroxisomal enoyl-CoA hydratase"
/protein_id="CAA76876.1"
/db_xref="GI:335325"
/db_xref="SPTREMBL:O74188"
/translation="VLASLSKMFRTAGLDLNEASNLVSETADMKEDGARTSLSMHKLI
RRQEARLTPDQAPFPVIAAVHGPVPFGLGVDLISSCDINIAHENSIFALKREYDVLAA
DITSLAYLPKITGNQSLIRELAYTGHAFSAEAKGLVSRILIPGSRREVVKAALELA
KTAGKSPIAVSGSKRLISHARDHSVSENLYVTQTWNAHAIVTKDMAECATSMKEKRT
PEFAPLKVPSKL"
3141..3199
intron
/gene="echA"
/number=1
3200..3435
exon
/gene="echA"
/number=2
3436..3510
intron
/gene="echA"
/number=2
3511..3840
exon
/gene="echA"
/number=3
3841..3898
intron
/gene="echA"
/number=3
3899..>3979
exon
/gene="echA"
/number=4
3968..3976
misc_feature
/gene="echA"
/number=4
/number=4
/note="peroxisomal targeting sequence"
BASE COUNT 1049 a 971 c 955 g 1183 t
ORIGIN

Query Match 1.68; Score 24; DB 8; Length 4158;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 actatgcaggtggcgcaaga 385
|||||
Db 1048 ACTATGCAGGTGGCGCAAGA 1071

RESULT 3
AP003048 157201 bp DNA linear PLN 22-MAY-2001
LOCUS
DEFINITION
ACCESSION AP003048
VERSION
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (cultivar:Nipponbare) DNA, clone:P0710A02.
Oryza sativa
Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartodeae; Oryzaceae; Oryza.
1 (sites)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0710A02
published Only in DataBase (2000) In press
2 (bases 1 to 157201)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Direct Submission
Submitted (13-DEC-2000) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program; Rannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
```

(E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468)
On May 21, 2001 this sequence version replaced gi:11862978.
Genes were predicted from the integrated results of the following:
GENSCAN1.0, BLASTx2.0, BLASTx2.0 as well as SplicePredictor
(October 1998 version). The genomic sequence was searched against
NCBI Nonredundant protein database, nr
(http://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
RGP. Protein homologues of the coding regions were searched against
NCBI Nonredundant protein database with BLASTx2.0. ESTs represent
the identified cDNA sequences using BLASTN 2.0 with the
corresponding dbj accession no. and RGP clone ID.
A gene with identity or significant homology to a protein is
classified based on the protein name to indicate the homology level
such as same name, 'putative-' and '-like protein'. A gene without
significant homology to any protein but with EST homology (covering
almost the entire length of partial sequence) is classified as an
'unknown' protein. A gene predicted with a gene prediction program
is classified as a 'hypothetical' protein.
The orientation of the sequence is from SP6 to T7 of the PAC clone.
Detailed information on overlap and assembly quality together with
annotation of this entry is available at
http://rgp.dna.affrc.go.jp/Genomeseq.html.
Location/Qualifiers
1. .157201
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="P0710A02"
join(4865..5500,5607..5668,5918..6014,6862..7143,
7382..7435,7532..7678)
/gene="P0710A02.1"
join(4865..5500,5607..5668,5918..6014,6862..7143,
7382..7435,7532..7678)
/gene="P0710A02.1"
/notes="contains ESTs AU065254 (E60905), AU065254 (E60905)
unknown protein"
/codon_start=1
/protein_id="BAB55803.1"
/db_xref="GI:14164404"
/translation="MPSDDEVREAGEVMILHFPPEAKSDNLFDKIHKNPASPAHFV
GRQEKARSKCTGDKVSIFQDGAQSEIAERGRDHDAEKSSKKDYTEAEAEF
LEVRILHRTFQKFNLLKDMVPHLSIDGPIDVANTENELRESRHGLDDPSMGFL
TERSPEPTNNQONPTLTENAYMTSECDYSGSPDHQEMNRESYGVKTSARA
GVFFCOPTASEEEMRIIPVFESSKKVLSFOLDPELTPEHLIKHLIAHESGHAI
VALHTQGAHPHQATVLPGLSLGVTVQVSPGNTSISKOKILARIDVCRGVAEEL
LFGHTVAGANDLFTAKELAKLVLLKLGESRERVKRLKKHEMQLQAVANALKEK
EILNDAEIRVQPTREPTQIFQKELAVNWS
complement(join(10619..10756,10904..11062,12001..12496,
12591..12733,13399..13580,13819..14023,14428..14565))
/gene="P0710A02.2"
complement(join(10619..10756,10904..11062,12001..12496,
12591..12733,13399..13580,13819..14023,14428..14565))
/gene="P0710A02.2"
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAB55804.1"
/db_xref="GI:14164405"
/translation="MIKWLIDIPQRTMKQSVLANHQPAFCGVNSYDHHKKRDERYLL
GMEISKWLATYIRNAEHTAYEFVAGSYDQAQLRLPLRFKTSYHCKIKSICY
DHVELYLTSTAGINNEPNTQIIFSLKSSNESYPIVYGTFAVDELDRLRN
YVFNTRNPVLEIOFINRDLRSNKKYLEVPVPHGTKSFISVESNNAHPIQDPTIG
NDLVPTPCRGMYILSFALFEDLVWKAESGGSDKQLLSAYAEINRGLKDPWYIR
IFSDBCLDVMCMFLNSVEAVIOVETKDDSDNPHRYFTAFSSGDEHETLVLDRLC
KKGKLFQNVAVKSEKLVIRLEFGSTFWTFODGAAVAVSSPDOSVSKLPDLWKA
RNDLKFQKLEPTQVCEAKMIKVTNELSNLMQEDLDEEDTNRMGGAASISTG
NGDCQILEVHFLTDSAITARTLAREEFKKNQGLCA"
join(18601..18688,19025..19173,19275..19365,19901..20054,
20134..20182,20516..20630,20751..20842,20938..21085,
21515..21579,21654..21751,22315..22420,22959..23096)
/gene="P0710A02.3"
join(18601..18688,19025..19173,19275..19365,19901..20054,

20134..20182,20516..20630,20751..20842,20938..21085,
21515..21579,21654..21751,22315..22420,22959..23096)
/gene="P0710A02.3"
/codon_start=1
/product="putative protein kinase AFC3"
/protein_id="BAB55805.1"
/db_xref="GI:14164406"
/translation="MPPEAATIGGSNLSAILSWSSPDALKSGFLSLSLSPLSPL
SPINESMESTSRKRARDQMDGAPPPAERAVVARGASPPWDDDRDDGHVFDL
GENNRCTGRVLECDRETHYVAIKVYRSIRKYDAMIDEIVDLRLAENRYRS
LCVQIQMFYRNHCICIVFKLGPLSYDLKRNRYQFPFVELVREFRQLLEAVYMH
ELRLHTDLKPNITLVSSSEYIRVPGSKNSQDEMHFKCLPKSSAIKLIDFGSTAFDM
QEHSSIVSTRHYRAPEIILGLGWSFPCDLMSVGCIIIVELCSGALPQTHENLEHLAMM
ERVGLPLPEHMRKASSAOKYFRGARLWNPGEAVSRRESIRAVKKDLRLKDLVARKA
DHSRAVLADLLYGLLKEPSERLTAOPLDHPFRNTT"
join(25678..25783,32101..32243,32405..32878,33109..33282,
33389..33526,33648..33785,33910..33971,34069..34342)
/gene="P0710A02.4"
join(25678..25783,32101..32243,32405..32878,33109..33282,
33389..33526,33648..33785,33910..33971,34069..34342)
/gene="P0710A02.4"
/notes="contains ESTs D24173(R1467), AU057957(S20436)"
/codon_start=1
/product="cytosolic aldehyde dehydrogenase"
/protein_id="BAB55806.1"
/db_xref="GI:14164407"
/translation="MAAANGDSKGEVPEKLEIKTKLEINGRFVDAVSGKTFETRPD
RTGEYIAIEGDKADIDILAVKAREAFDHPWPRMSGFARGRLHKFPADLVEHVEE
LAALDTAGDKLFGAMGLVDIPGGANLLRYIAGADKDVHGETLKKMPCRGCTKEPV
GVGHIIPWNYPTTFMFFKSPALAACTMVKVPAPBOTPLSALFYAHLAKLAGPDGV
LNVYFGPTAGAAISSHMDIDKVSSTGEYGRLYVMEAAKSNLKPVSLELGGKSPV
IVFDADLDTAVNLNHSYNTNKEIGCVAGSRIYVQEQYIDAFVKATMAKKSVMGD
PFNPRVHQGPQIOKEQYKILYIDIGKREGATLVGTGKPCGSGYIEPTITFDVKE
EMSAOBEIFGPVMAIMKFTVEAKNSTRYGLAAGITVKNDIVANTVRSIRAG
AINTNCLYGFDPVPGCYGKMSGFGKMGMDALEKYLHTKAVVTPLYNTPEWL"
complement(39757..39978)
/gene="P0710A02.5"
complement(39757..39978)
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAB55807.1"
/db_xref="GI:14164408"
/translation="MHRKEIINHNEIDGNDVVEAKLVIPGMAERDKMRLDNV
GLVLRGSGATNDMPDARGDORRSATARRL"
join(42079..42199,45522..45664,45785..46168,46289..46378,
46469..46642,46767..46904,47043..47180,47253..47314,
47427..47700)
/gene="P0710A02.6"
join(42079..42199,45522..45664,45785..46168,46289..46378,
46469..46642,46767..46904,47043..47180,47253..47314,
47427..47700)
/gene="P0710A02.6"
/notes="contains EST AU164502(R0133)"
/codon_start=1
/product="putative cytosolic aldehyde dehydrogenase"
/protein_id="BAB55808.1"
/db_xref="GI:14164409"
/translation="MGSTGCGNGKAAAGCGGLVPEIETFKLPIFINGERVDAASKTF
KTRDPRTGDLVLAIAEADKADVDLAKAAREAFHCKWPRMSGERSRVRMKLADLVE
QHADEALADGADAGKLLTGKIIDMPAAQNMRYAGAADKHGESLRVAKYQGYT
LREPIGVGVIIIPWNTFMFFLKVPALAACTIVVKAQETPISALYIAHLAKLAG
VPDGVINVYFGPTAGAAISSHMDVDSVAFTGSAEIGRAIMESAASNLKNVSLLEL
GKSPMIVFDADVDMAVLSLAVFENKGEICVAGSRVTVQEQYIDFVFKKAAEAKN
WKGVDPMDDATNMGPVDKVOFERVLKYIEIGKNEGATLLTGKGTGDKGYIETPIF
VDKEETIAQERIFGPVMSLMKFTVEAKNCTKYGLAAGITVKNLNTANVSR
SVRAGTVMVCFAPDAPPGCYKMSGFGCGGCGOOWAMDKYLOVKTIVTAVPDSFWY"
complement(52056..52268)
/gene="P0710A02.7"
complement(52056..52268)
/notes="hypothetical protein"
/codon_start=1


```

/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative Trm2011-2 protein"
/protein_id="CAC48695.1"
/db_xref="GI:15140168"
/translation="MGAELVQDDVAGFEFRHONLLDIDAFAIDRAIEDKGSADAV
GPCGDKHGVPMAEGVVAHQSFARAPAPORRHVRRLRPGLIDENEAGGINSAGVLP
SCPVAGDVGLPFPAGDQRLFL"
/complement(1301..1708)
/gene="Trm2011-2N OR Smb20305"
/complement(1301..1708)
/gene="Trm2011-2N OR Smb20305"
/function="ELEMENTS OF EXTERNAL ORIGIN; Transposon-related
functions"
/notes="Product confidence : probable
Gene name confidence : putative"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="probable ISRM2011-2 transposase protein,
N-terminal portion"
/protein_id="CAC48696.1"
/db_xref="GI:15140169"
/translation="MARPSNDLRRVVDAVTGELSCRAAKRFGIGISTADWVR
FRFGSAAPQMGHKKPKLSGPHAWLLCRERDFTLHGLVAELSERGLKVDYRAV
WTVHREGLSYKRRWSPANGSGPTSPATGRDG"
/complement(1801..2448)
/gene="Smb20307"
/complement(1801..2448)
/gene="Smb20307"
/function="MISCELLANEOUS; Hypothetical/Partial homology"
/notes="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative dihydroxyacetone kinase protein"
/protein_id="CAC48697.1"
/db_xref="GI:15140170"
/translation="MMSESAVRRLIEVFATISSAIEERDLSELDAIGDADHGIMT
AIGDSVTELTARTAGSAGLVMTAAASFLNAVGAATGPLYTAFRAAAQALGR
DITDIEACSLLIQVAVSGIGERGKQGDKTMLDVLPADAQAQVAAAKSTAFAWA
DVTAAANGASATRSMTVKGAARVGRSLGHMDPGGAASAVLIIRAMAQTLREG"
/complement(2527..4614)
/gene="Smb20312"
/complement(2527..4614)
/gene="Smb20312"
/function="MISCELLANEOUS; Not classified regulator"
/notes="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcriptional regulator protein"
/protein_id="CAC48698.1"
/db_xref="GI:15140171"
/translation="MTIKPSARKGAANRSCTIDASDAIPRYGDDPYVWACMLYED
GMTQIEACMGVSRATVNSYLADAREGIWNISIEFARLASLTIAQELRHFGLTDC
LVVSDNNARLIDRLGAGAOALPKLLSGDTLAVAGRTVLSVGHAGIASLQDMT
VVOATGGTTAPAYTPELCSAVARAI SARCVNTAPAVKSPELLHMLEPLVQEQ
FATLARANVLEGISLSRPNSTIHTSGFESVLDQYLAAGAVGVAGRTIDGRPV
AGPLDRTVGISLIDLLRKITRIAVAGCFDKVPALLAALRGVYVNLITDAATGHGL
NGSHVALVARNRPKVLVPTPSSYRTHVKFLNPNVEEMLDGAVKAHETYLQPI
VQCAASGEGVLFVYCNAGDVNFEAAEIAEEKIKVKVTLITDDVASSPDDR
EGRRVAGHFFTEKTAGACDRGLSLEACEAVTRKANLRTYTVGVALPGRSNFQTRRH
NEIGPDDMEVGMGIGHGEGVTRERIRSADEITSDIMDRIFKEMKTRPGRSVAVLNS
FATPLWELTILFRKQVQRLAAKDLIIIEANNWIGHYCTSLDMAGASITILHLDQELSL
LHPCTETAFKVN"
/complement(4728..5369)
/gene="Smb20313"
/complement(4728..5369)

/gene="Smb20313"
/function="Small Molecule Metabolism; Central intermediary
metabolism; misc. glycerol metabolism"
/notes="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative dihydroxyacetone kinase protein"
/protein_id="CAC48699.1"
/db_xref="GI:15140172"
/translation="MOTFENARAGDIVLTADRIENRAYLSIDGKIGDGHGVNMA
KGFGLAARLKGDSALAEALDTLGTVMTEIGSGMCPLYGVMTFEAFKIEGADAID
AAAYSRMLHAGLEGIRISGSAKVGDKTLDTLPVDAFDAATADGRPAEALDALVS
AAEAGRDSTROLIARIGRASHRGLGSLVDAGATSCAIIILRELSHAGARALA"
/complement(5380..6372)
/gene="Smb20314"
/complement(5380..6372)
/gene="Smb20314"
/function="Small Molecule Metabolism; Central intermediary
metabolism; misc. glycerol metabolism"
/notes="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence=not_experimental

Query Match 1.5%; Score 22; DB 1; Length 260050;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 gccgttcggtcagcgagcgatca 184
|||||
Db 175408 GCGGCTTCGGTCAGCGGCTCA 175429

RESULT 5
HSU62769 3897 bp mRNA linear PRI 15-AUG-1997
LOCUS Human oxytocinase variant 2 mRNA, complete cds.
DEFINITION
ACCESSION U62769
VERSION U62769.1 GI:2209277
KEYWORDS
SOURCE human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 3897)
Laustsen, P.G., Rasmussen, T.E., Petersen, K., Pedraza-Diaz, S.,
Moestrup, S.K., Gliemann, J., Sottrup-Jensen, L. and Kristensen, T.
The complete amino acid sequence of human placental oxytocinase
Biochim. Biophys. Acta 1352 (1), 1-7 (1997)
97320624
REFERENCE
2 (bases 1 to 3897)
Laustsen, P.G., Rasmussen, T.E., Petersen, K., Moestrup, S.,
Gliemann, J., Sottrup-Jensen, L. and Kristensen, T.
Direct Submission
Submitted (02-Jul-1996) Department of Molecular and Structural
Biology, Aarhus University, Langelandsgade 140, Aarhus 8000 C,
Denmark
FEATURES
source
1. 3897
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="placenta"
359..3394
/function="inactivates oxytocin and vasopressin"
/notes="cysteine aminopeptidase; vasopressinase; leucine
aminopeptidase"
/codon_start=1
/product="oxytocinase splice variant 2"
/protein_id="AAB66573.1"
/db_xref="GI:2209278"

```

```
/translation="MIENSMFEPEPDDVDLAKEPCLHPLPEDEVEYPRGSRLLVRGL
GRHEDEEDYESSAKLLGMSFMRSSGLRNSATGYRQSPDGACSVPSARTMVWCAF
VIVAVSVIMVYILLPCTFTKECHKKNOSIGLIOFPATNGKLFPAQIRLPTAVVP
LEYELSHPNLTSMTFRGSVTVISQALOVWVNIILHSTGHNISRVTFMBSAYSSQKOA
ELLETAYHGQIAIYAPCALAGHNTLTIKIETSANISSYIGFYGFSTDDSEKKTYFA
ATQCFALARSAPCFDEPAFAKTFILKIINDEQYITALSNMKPKSSVLDGGLVQDEF
SSVKMSYLVAFIVGEMKNLSQDVNGTLTVSYAVPENICOVHYALETVTKLLEFFON
YFEIQYPLKLDLVAIPDFEAGAMENGLLTFREETLLYDSNTSSMADRKLVTKIIAH
ELAHQWFLNLTVMKWNDLWNEGFAFMFYFSLKIKFELKLSYEDVFOAHVLYLHNHS
SLNSSHPISSVSQSEQUEEMFDSLSYFKGSSLLMLKTYLSEDEVFOHVVLYLHNHS
YASIODDLDWSENEVTNOTLDVVRMKTWTLOKGRPLVTQVKKGELFTQOERRFLN
MKPEIOPDSTVYLWHIPLSYVTEGNTSKYSLSLQKSSVJNLPERVLWKVNIIM
NGYVIVHADVADDEALHOLKINBYVLSDDKDRANLINNIELAGLKVBLKRAFDLIN
YLGNNHTAPTEALFQTDLLYNLEKLYGMDLASRLVTVRVFKLLQVQQOQVWIDEG
TPSMELNSALLEFACHTNLNCSSTAKULFDMMASNGTOSLPTDQVMTVTVFKVAGK
DKGWSFLGKYSIGSSEAENKILFALASSEDVVKLYWLMKSSLNGDNNFTOKLSII
RTVGRRHFCHLLANDFVKENNKLVQKPELGSYTIQIVAGSYTLFSTKTHLSEVOAF
FENQSEATFLRLCVOEALEVIQLNIQWENKLNKLSLTMWL"
BASE COUNT      1142 a  787 c  846 g  1122 t
ORIGIN

Query Match      1.4%; Score 21; DB 9; Length 3897;
Best Local Similarity 100.0%; Pred. No. 7.4; Mismatches 0; Gaps 0;
Matches 21; Conservative 0; Indels 0;

Qy 525 ctgctcttcattcttcaaacg 545
|||||
Db 307 CTGCTCTTCATCTTCAAAGC 327

RESULT 6
HSA131024      4226 bp  DNA  linear  PRI 15-JUN-2000
LOCUS Homo sapiens otase gene, exon 1-alpha.
DEFINITION
ACCESSION AJ131024
VERSION AJ131024.1 GI:6468557
KEYWORDS otase gene; oxytocinase/insulin-responsive aminopeptidase.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4226)
Rasmussen, T.E., Pedraza-Diaz, S., Hardre, R., Laustsen, P.G.,
Carroll, A.G. and Kristensen, T.
Structure of the human oxytocinase/insulin-regulated aminopeptidase
gene and localization to chromosome 5q21
Eur. J. Biochem. 267 (8), 2297-2306 (2000)
2023264
REFERENCE 2 (bases 1 to 4226)
Rasmussen, T.E.
Direct Submission
JOURNAL Submitted (23-NOV-1998) Rasmussen T.E., Department of Molecular and
Structural Biology, University of Aarhus, DK-8000 Aarhus C,
Jylland, DENMARK
FEATURES
source
1. 4226
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q21"
295..546
/rpt_family="Alu"
/rpt_unit=295..546
/rpt_type=DISPERSED
1275..1470
/rpt_family="Alu"
/rpt_unit=1275..1470
/rpt_type=DISPERSED
3412..3737
/gene="otase"
/note="number 1-alpha"
/usedin-AJ131023:otase_mRNA
```

```
/usedin-AJ131023:otase_CDS
3412..3737
/gene="otase"
3809..3990
/rpt_family="Alu"
/rpt_unit=3809..3990
/rpt_type=DISPERSED
BASE COUNT      1093 a  824 c  831 g  1466 t  12 others
ORIGIN

Query Match      1.4%; Score 21; DB 9; Length 4226;
Best Local Similarity 100.0%; Pred. No. 7.5; Mismatches 0; Gaps 0;
Matches 21; Conservative 0; Indels 0;

Qy 525 ctgctcttcattcttcaaacg 545
|||||
Db 3709 CTGCTCTTCATCTTCAAAGC 3729

RESULT 7
AB008831      10498 bp  DNA  linear  BCT 05-JUN-1998
LOCUS Acinetobacter sp. genes for aniline dioxygenase subunits and
DEFINITION reductase component, partial and complete cds.
ACCESSION AB008831
VERSION AB008831.1 GI:2627146
KEYWORDS 4-oxalocrotonate decarboxylase; 4-hydroxy-2-oxovalerate aldolase;
acetaldehyde dehydrogenase; 2-oxopent-4-dienoate hydratase;
2-hydroxyisocaproate semialdehyde hydrolase; 2-hydroxymuconic
semialdehyde dehydrogenase; catechol 2,3-dioxygenase; small
ferredoxin-like protein; aniline dioxygenase reductase component;
aniline dioxygenase beta-subunit; aniline dioxygenase
alpha-subunit.
SOURCE Acinetobacter sp. (strain:YAA) DNA.
ORGANISM Acinetobacter sp.
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Acinetobacter.
REFERENCE 1 (bases 1 to 10498)
Takeo, M.
Direct Submission
JOURNAL Submitted (13-NOV-1997) Masahiro Takeo, Himeji Institute of
Technology, Department of Applied Chemistry; 2167 Shosha, Himeji,
Hyogo 671-22, Japan (E-mail:takeo@chem.eng.himeji-tech.ac.jp,
Tel:0792-67-4893, Fax:0792-67-4891)
2 (sites)
Takeo, M., Fujii, T., Takenaka, K. and Maeda, Y.
Cloning and sequencing of a gene cluster for the meta-cleavage
pathway of aniline degradation in Acinetobacter sp. strain YAA
J. Ferment. Bioeng. 85, 514-517 (1998)
3 (sites)
Fujii, T., Takeo, M. and Maeda, Y.
Plasmid-encoded genes specifying aniline oxidation from
Acinetobacter sp. strain YAA
Acetobiology (Reading, England) 143 (Pt 1), 93-99 (1997)
JOURNAL 9717776
MEDLINE
PUBMED 9025282
FEATURES
source
1. 10498
/organism="Acinetobacter sp."
/strain="YAA"
/db_xref="taxon:472"
1..276
/gene="atdA3"
<1..276
/gene="atdA3"
/codon_start=1
/transl_table=11
/product="aniline dioxygenase alpha-subunit"
/protein_id="BA23550.1"
/db_xref="GI:2627147"
/translation="ELNAIRMRTOEDFPIMGEYDDVANFESCOEGLETWPEIWDIFS
RHMNEGENACYQDVIIQHKPTSEIHSRHYFTWLQLMSAVNKQNSV"
```

| | | | | | |
|------|------------|--|------------|---|------------|
| gene | 278..874 | /gene="atd4" | 4187..5647 | /gene="atdc" | 1487..5647 |
| CDS | 278..874 | /codon_start=1 | 4187..5647 | /codon_start=1 | 1487..5647 |
| | | /gene="atd4" | | /transl_table=1 | |
| | | /codon_start=1 | | /product="2-hydroxymuconic semialdehyde dehydrogenase" | |
| | | /transl_table=1 | | /protein_id="BAA23556.1" | |
| | | /product="aniline dioxygenase beta-subunit" | | /db_xref="GI:2627153" | |
| | | /protein_id="BAA23551.1" | | /translation="MKDKIKHINQYVGSASGKLFNNVNPANGKLSKYHEAGREYD | |
| | | /db_xref="GI:2627148" | | AADNKAALRPGMGMKMTLEERTSILHKVADGINAREDEFLKAECLDTGPKMSASHI | |
| | | /translation="MNNKDSIKIVPSAVNLAAYQEIKOYDYFWNLGNLGEALD | | DIPGAARFKVADMIKINYESEFTPDGTALNAVVRPKGVIGVISPWNPLLL | |
| | | HKINFLKEARLDQOCQDEMLFLEDGCIYWPGPSAPASEATYFHDRIIRLK | | MTWKVPALACANVVKVPSDETPTTALLGEVNMKEAGIPTVYVYVHGFGSAGAF | |
| | | DRIVLQCFAYSQIPVSKNRIILGAPEVWAPVGSSECFVLTSTFIVFESRDGKSOVL | | LTEHPQVDFTFGTGTGPEVVKAAKGRDISELGGKNAALVYFADCDMDKAIIEGT | |
| gene | 885..1895 | SGWYGVITIKONDELKIKKQINLNDCLSPQGNNSFFL" | | MRSAFANGGVCLGTERTVYVERSIYDFEVLGRLEAAEGLKIGPPDDAENLPLVSLK | |
| CDS | 885..1895 | /gene="atd4" | | HREKVLSTYQKAVDDGATVVTGGGVPEMPELAGANVQPTIWTGLPKDSAVITDRIIP | |
| | | /gene="atd4" | | GPCVHVCPPDTEEAIELANSLPYGLASAIWSENITRAHRAVQAGIENGIIWVNSWELR | |
| | | /codon_start=1 | | DLRTPFGSKSGIGREGGVHSLFETEMKNICVKL" | |
| | | /transl_table=1 | | 5655..6512 | |
| | | /product="aniline dioxygenase reductase component" | | /gene="atdb" | |
| | | /protein_id="BAA23552.1" | | 5655..6512 | |
| | | /db_xref="GI:2627149" | | /gene="atdb" | |
| | | /translation="MNTLFRVIDIKIETKESRFSVLKPLDGLVLAHSPGKYLPIKIR | | /codon_start=1 | |
| | | TKGGLLFRSYSLSSASANEDFKITVRREGSGSNMCDNVKVGDIETLPPAGSFH | | /transl_table=1 | |
| | | POMWRDQVFPAGSGITPVISIITKALNRHKNRIKLFYANSSSESIIFHKLKDLCL | | /product="2-hydroxymuconic semialdehyde hydrolase" | |
| | | QPPRLDQFWDDEKGIPTSIATFEQITDDALEVEYFLCGPAPFMGVENFLKESVP | | /protein_id="BAA23557.1" | |
| | | PGLIITKESFAGSVDDNGDVTVESSAEDKVTNFMNLGINKSMVCSDDFTLNEIIRAG | | /db_xref="GI:2627154" | |
| | | INVPSSCCAGNCGSMCLLVSGDVIETNSVTLDASDEEDGWILACRSKPKRSKNIETSF | | /translation="MTAPONSPEIGREITAGYRTVNVHDHGADNADKSDVAVMHIHSGG | |
| | | DO" | | PGVTAWANRMLVPELAKNRRLVAPDMLGFGYTERPEDNTYNNRERWYKHAIVGMDELG | |
| gene | 1902..2828 | /gene="atdr" | | LEQYDLVNSFGGLALALAEIHPQRRRLVLMGVSVPFTITKGLDEVGMGYESIENM | |
| CDS | 1902..2828 | /gene="atdr" | | RLMDVFAYNNKLLTDELAEMRYQASVYRPGFQESFAAMFAPRQPLDNLNLSPEEDIR | |
| | | /note="similar to LysR-type transcriptional activator" | | KLPHEILIHGREDVITPLEVLQLAELIDRAHNSNVFCRCATITQIEHAGRFARLVNDF | |
| | | /codon_start=1 | | LHEADOPPE" | |
| | | /transl_table=1 | | 6516..7184 | |
| | | /protein_id="BAA23553.1" | | /gene="atdb" | |
| | | /db_xref="GI:2627150" | | 6516..7184 | |
| | | /translation="MRCVIRKSSSYDNANDNVKAFALYRARKDYEAASLESGLVDGST | | /gene="atdb" | |
| | | LRKKTSQEVLYGSLFVKENNRWALAPGDEYDLKAAVDMESATRLFFGLPDGKGV | | /codon_start=1 | |
| | | IKVSPVLYNQFSDVLIARFIYPTFOISSDAVEDLEREGFDFIARLAPIGNM | | /transl_table=1 | |
| | | NSLKIKRGTGKLVGFGSASYDDIVDFPGKDALRDKSLIKTGIGFSYKPHDFVFLG | | /product="2-oxopent-4-dienoate hydratase" | |
| | | LMQELQFVGNIKTMCODLESCARFCDKGAGLALPFLAHRYPNNSCVYDASDTLMA | | /protein_id="BAA23558.1" | |
| | | ELWLISRLDKSDQVVLGNLMSAKSKEVMS" | | /db_xref="GI:2627155" | |
| gene | 2888..3226 | /gene="atds" | | /translation="MEQAKIQAFADLEYALVNRKAVPLTSAODITTEDAYHISLR | |
| CDS | 2888..3226 | /gene="atds" | | MLERQOAGARIIGKKTGVTSKAMNMLNVHQPFGVLTDDMVFNSEGVNIDRLIA | |
| | | /codon_start=1 | | PRAGETAFLLKDLTGCTVNTADVLATCEVMPCEIVDSRIKDKKIAIOTDIAINA | |
| | | /transl_table=1 | | SCGLFVLCDKAVSPREVDLVTCGVNVEKNGAILSTGAGAAALGSPVNCVTLANTLQO | |
| | | /product="small ferredoxin-like protein" | | FGIR" | |
| | | /protein_id="BAA23554.1" | | 7290..8198 | |
| | | /db_xref="GI:2627151" | | /gene="atdb" | |
| | | /translation="MGESYQITEQCSQRPFCRQAGSVLKAMEQQGLECAPVCRCGGG | | 7290..8198 | |
| | | CGJCVTVREGDYFCGKMSRVHAPPEALAQGEVLACRIYPLSDLIIECRPROSAAGLA | | /gene="atdb" | |
| | | NEYTTTKAMR" | | /codon_start=1 | |
| | | /gene="atdb" | | /transl_table=1 | |
| | | 3232..4155 | | /product="4-hydroxy-2-oxovalerate aldolase" | |
| | | /gene="atdb" | | /protein_id="BAA23560.1" | |
| | | 3232..4155 | | /db_xref="GI:2627157" | |
| | | /gene="atdb" | | /translation="MNLGKKVTLHDMSLPDGMHAKRHOISLDEMYSIATGMDACMP | |
| | | /codon_start=1 | | LIEVTHGDLGGRSLNCFPAHSDDEYLSAVYPMKNKAKISALLLPGITGVYDHLKNAK | |
| | | /transl_table=1 | | DLGVTIRVATHTCTEADVSOQHIGMAKMGMDTVGFLMAHNVSPENPGAKUMVGTG | |
| | | /product="catechol 2,3-dioxygenase" | | ANCYATDSAGYMLPDEVYAKIIGLLRSELPSNIEVFGFHGHHNMGAIANSLAARSGAS | |
| | | /protein_id="BAA23555.1" | | | |
| | | /db_xref="GI:2627152" | | | |
| | | /translation="MKKGVMRPGHVQIVRLDMDAEVAKHYKDLGLIEMDRDQGRVYL | | | |
| | | KAWTVDFSVLREADEPFGMDPFGKCLSEEVVDQLRGELAAFGCEIYDEPELDK | | | |
| | | CCRRVFTAPYTHGHSFEIYATKQTKGKGVGNHNPAMPRLGCMKATRFDDHCLLYPN | | | |
| | | LDLNTLFRDVLGDLAEQVMAQPKRVSOFLTVYSTKAHDIAFIHHEEPKTHHASF | | | |
| | | LETWQDVLKAADLLSMYDTSIDIGPTRHGLTHGTIYFFDPSGNRNEVAGGDIHYPD | | | |
| | | HEPTWDAEELGKAIFYHDRVLRNERFWYLT" | | | |
| gene | 4187..5647 | /gene="atdc" | | | |

RIDGSAVLGAGAGNTPLFVFCALVDRMGVETGVLDLYKIMDVAEEDLVVPIMDQPIRVD
RNALTLGAGVYSSSELLFAEAKKYGISARDILVELGRGTGCGQEDMEDIALITMA
KEKGLI*

Query Match 1.4%; Score 21; DB 1; Length 10498;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1201 acggaagaagagccattgag 1221
|||||
Db 5393 ACGGAAGAAGAGGCCATTGAG 5413
|||||

RESULT 8
AL356308 102910 bp DNA linear PRI 14-DEC-2001
LOCUS Human DNA sequence from clone RP11-475L16 on chromosome 10,
DEFINITION complete sequence.
ACCESSION AL356308
VERSION AL356308.16 GI:17894610
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
Tracey, A.
Submitted (14-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Dec 17, 2001 this sequence version replaced gi:16972876.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-475L16 is from the library RPCI-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone
RP11-475L16. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true right end of clone RP11-475L16 is at 102910 in this
sequence. The true left end of clone RP11-557K21 is at 17539 in
this sequence. The true right end of clone RP11-146P20 is at 2000
in this sequence.
FEATURES
Location/Qualifiers
1. 102910
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-475L16"
/clone_lib="RPCI-11.2"
BASE COUNT 32050 a 20754 c 19228 g 30878 t

ORIGIN

Query Match 1.4%; Score 21; DB 9; Length 102910;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 830 cggcaagagattctcaaaat 850
|||||
Db 72046 CGGCAAGGATTCTCAAAAT 72066
|||||

RESULT 9
AC008850/c 131805 bp DNA linear PRI 25-SEP-2001
LOCUS Homo sapiens chromosome 5 clone CTD-217216, complete sequence.
DEFINITION AC008850
ACCESSION AC008850
VERSION AC008850.7 GI:15741477
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 131805)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 131805)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 3 (bases 1 to 131805)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Sep 25, 2001 this sequence version replaced gi:7711359.
Draft Sequence produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.5% of Sequence;
Estimated Total Number of Errors is 0.5.

FEATURES
source
1. 131805
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-217216"
BASE COUNT 43916 a 26019 c 24296 g 37574 t

ORIGIN

Query Match 1.4%; Score 21; DB 9; Length 131805;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 525 ctgctcttcattctcaaacg 545
|||||
Db 88638 CTGCTCTTCATCTCAAGC 88618
|||||

RESULT 10
AP003928/c 138904 bp DNA linear HTG 19-JUL-2001
LOCUS AP003928
DEFINITION Oryza sativa chromosome 8 clone OJ1150_A11, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION AP003928
VERSION AP003928.1 GI:14915711
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Oryza sativa (cultivar:Nipponbare) DNA, clone:OJ1150_A11.
ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE AUTHORS

1 (bases 1 to 138904)
Oryza sativa nipponbare(T. and Yamamoto, K.
clone:Oul150_All

JOURNAL REFERENCE

2 (bases 1 to 138904)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission

JOURNAL TITLE

Submitted (18-JUL-2001) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: tsasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,
Tel: 81-298-38-7441, Fax: 81-298-38-7468)

COMMENT

The nucleotide sequence of this BAC clone was generated by
combining Monsanto and RGP-Japan sequencing data.
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES source

Location/Qualifiers
1..138904
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="8"
/clone="Oul150_All"
BASE COUNT 40002 a 28323 c 28941 g 41170 t 468 others
ORIGIN

Query Match 1.4%; Score 21; DB 2; Length 138904;
Best Local Similarity 100.0%; Pred. No. 8.8;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 510 gccctcgccgcggcgctc 530
|||||
Db 47643 GCCCTCGCCGCGCTC 47623

RESULT 11

AC008907 147344 bp DNA linear HTG 20-APR-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2261G7, WORKING DRAFT SEQUENCE,
24 unordered pieces.
AC008907
AC008907.5 GI:13699452
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEPIN.
KEYWORDS human.
SOURCE

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 147344)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 20, 2001 this sequence version replaced gi:7709395.
-----Genome Center

COMMENT

Center: Joint Genome Institute
Center Code: JGI

Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 727924
Center clone name: CITB-HI_2261G7

Summary Statistics
Consensus quality: 122801 bases at least Q40
Consensus quality: 131739 bases at least Q30
Consensus quality: 134984 bases at least Q20
Estimated insert size: 123000; pulse-field gel estimation
Estimated insert size: 145044; sum-of-contigs estimation
Quality coverage: 8.83 in Q20 bases; pulse-field gel estimation
Quality coverage: 7.49 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1..1024: contig of 1024 bp in length
1025 1124: gap of unknown length
1125 2152: contig of 1028 bp in length
2153 2252: gap of unknown length
2253 3267: contig of 1015 bp in length
3268 3367: gap of unknown length
3368 4464: contig of 1097 bp in length
4465 4564: gap of unknown length
4565 5581: contig of 1017 bp in length
5582 7285: contig of 1604 bp in length
7286 7386: contig of 1177 bp in length
7387 8662: gap of unknown length
8663 9732: contig of 1069 bp in length
9733 11051: contig of 1220 bp in length
11052 11151: gap of unknown length
11152 12313: contig of 1162 bp in length
12314 12412: contig of 1029 bp in length
12413 13542: gap of unknown length
13543 14997: contig of 1455 bp in length
14998 16437: contig of 1340 bp in length
16438 16538: gap of unknown length
16539 17545: contig of 1008 bp in length
17546 19312: contig of 1667 bp in length
19313 21084: contig of 1652 bp in length
21085 21164: gap of unknown length
21165 23214: contig of 2050 bp in length
23215 29721: contig of 6407 bp in length
29722 29821: gap of unknown length
29822 39205: contig of 9384 bp in length
39206 39305: gap of unknown length
39306 54491: contig of 15185 bp in length
54492 66531: contig of 11941 bp in length
66532 85895: contig of 19264 bp in length
85896 105738: contig of 19743 bp in length
105739 105838: gap of unknown length
105839 147344: contig of 41506 bp in length.

FEATURES

Location/Qualifiers
1..147344
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

/chromosome="5"
/clone="CTD-2261G7"
/clone_lib="Caltech human BAC library D"
BASE COUNT 42045 a 27306 c 28277 g 47403 t 2313 others
ORIGIN

Query Match 1.4%; Score 21; DB 2; Length 147344;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 525 ctgctcttcattctcaagc 545
|||||
Db 14097 CTGCTCTTCATCTTCARAGC 140117

RESULT 12
AL627302 191415 bp DNA linear HTG 30-JAN-2002
LOCUS Mus musculus chromosome X clone RP23-272D10, *** SEQUENCING IN
DEFINITION PROGRESS ***, in ordered pieces.
ACCESSION AL627302
VERSION AL627302.18 GI:18476832
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (sites)
Direct Submission
AUTHORS Bates, K
TITLE Submitted (29-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 1, 2002 this sequence version replaced gi:18250545.
COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BM272D10
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 9% of reads
Chemistry: Dye-primer Big Dye; 0% of reads
Consensus quality: 191248 bases at least Q40
Consensus quality: 191351 bases at least Q30
Insert size: 191415; sum-of-contigs
Insert size: 182914; 6.3% error; agarose-fp
Quality coverage: 13.20x in Q20 bases; sum-of-contigs Quality
coverage: 13.82x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1. .191415
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-272D10"
/clone_lib="RPCI-23"
misc_feature
1. .191415
/note="assembly_fragment:03353
clone_end:SP6
vector_side:left"
BASE COUNT 52805 a 40506 c 40442 g 57662 t
ORIGIN

Query Match 1.4%; Score 21; DB 2; Length 191415;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1125 tgtcaatccccaaaccaggcag 1145
|||||
Db 105060 TGTCAATCCCAACACGCGAG 105080

RESULT 13
AC008908 251664 bp DNA linear HTG 19-APR-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2265E4, WORKING DRAFT SEQUENCE,
DEFINITION 16 ordered pieces.
ACCESSION AC008908
VERSION AC008908.5 GI:7711373
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 251664)
Sequencing of Human Chromosome 5
Unpublished
REFERENCE 2 (bases 1 to 251664)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On May 6, 2000 this sequence version replaced gi:7709396.
COMMENT ----- Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 729409
Center clone name: CITB-HL_2265E4
-----
Summary Statistics
Consensus quality: 232454 bases at least Q40
Consensus quality: 246311 bases at least Q30
Consensus quality: 248647 bases at least Q20
Estimated insert size: 218000; pulse field gel estimation
Estimated insert size: 250914; sum-of-contigs estimation
Quality coverage: 5.4 in Q20 bases; pulse field gel estimation
Quality coverage: 5.56 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1
25893: contig of 25893 bp in length
25894 25993: gap of unknown length
25994 39631: contig of 13638 bp in length
39632 39731: gap of unknown length
39732 61054: contig of 21323 bp in length
61055 61154: gap of unknown length
61155 77299: contig of 16145 bp in length
77300 77399: gap of unknown length
77400 98305: contig of 20906 bp in length
98306 98405: gap of unknown length
98406 109130: contig of 10725 bp in length
109131 109230: gap of unknown length
109231 112039: contig of 2809 bp in length
112040 112139: gap of unknown length
112140 126605: contig of 14466 bp in length

```

* 126506 126705: gap of unknown length
 * 126706 132688: contig of 5983 bp in length
 * 132689 132788: gap of unknown length
 * 145777 145777: contig of 12989 bp in length
 * 145778 145877: gap of unknown length
 * 145878 150206: contig of 4329 bp in length
 * 150207 150306: gap of unknown length
 * 150307 180051: contig of 29745 bp in length
 * 180052 180151: gap of unknown length
 * 180152 208645: contig of 28494 bp in length
 * 208646 208745: gap of unknown length
 * 208746 217750: contig of 9005 bp in length
 * 217751 217850: gap of unknown length
 * 217851 237954: contig of 20104 bp in length
 * 237955 238054: gap of unknown length
 * 238055 251664: contig of 13610 bp in length.

FEATURES

source
 1. .251664

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone_lib="Calttech human BAC library D"
 /clone="CTD-226584"
 BASE COUNT 80411 a 47554 c 46337 g 75858 t 1504 Others
 ORIGIN

Query Match 1.4%; Score 21; DB 2; Length 251664;
 Best Local Similarity 100.0%; Pred. No. 9.1;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 525 ctgctctctcatcttcaaacg 545
 |||||
 Db 186250 CTGCTCTCTCATCTTCAAGC 186230

RESULT 14

FVNS5GAI
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 FEATURES

source

1. .481
 /organism="Sepik virus"
 /strain="MK7148"
 /db_xref="taxon:44026"

gene

1. .251

CDS

<1..251
 /gene="NS5"
 /codon_start=3
 /product="nonstructural protein"
 /protein_id="AA02061.1"
 /db_xref="GI:1066551"

translation="LXLSELATRPCLWTKPKISXGTTQASWARDVKDAGHRIRK
 LVNKEFYDYMSTMGRYRELGEGRFTLLDRLTQHP"
 BASE COUNT 133 a 122 c 127 g 87 t 12 Others
 ORIGIN

Query Match

1.4%; Score 20; DB 14; Length 481;

Best Local Similarity 100.0%; Pred. No. 25;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1390 agagagctaggcgagaagg 1409

|||||
 Db 186 AGAGAGCTAGCGGAGAGG 205

RESULT 15

AB009654
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE

1 (sites)
 Alteromonas sp. DNA for aldehyde dehydrogenase, complete cds.
 Alteromonas sp. (strain:KE10) DNA.
 Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;
 Maeda,T., Yoshinaga,I., Shiba,T., Murakami,M., Wada,A. and
 Ishida,Y.

TITLE

Cloning and sequencing of the gene encoding an aldehyde
 dehydrogenase that is induced by growing Alteromonas sp. Strain
 KE10 in a low concentration of organic nutrients
 Appl. Environ. Microbiol. 66 (5), 1883-1889 (2000)

JOURNAL

MEDLINE

AUTHORS

TITLE

JOURNAL

REFERENCE

FEATURES

source

location/Qualifiers

organism

strain

db_xref

note

gene

CDS

codon_start

transl_table

product

protein_id

db_xref

translation

SD

TW

IP

FG

DHE

DT

NH

GR

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Oy

Db

1.4%; Score 20; DB 1; Length 2001;

Best Local Similarity 100.0%; Pred. No. 27;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1201 acggaagaagagccattga 1220

|||||
 Db 1471 ACGGAAGAAGAGCCATTGA 1490

Search completed: April 27, 2002, 05:34:37
Job time: 17314 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:46:04 ; Search time 114.27 Seconds
(without alignments)
825.083 Million cell updates/sec

Title: US-09-882-694A-3
Perfect score: 2902
Sequence: 1 MSATNSRGDCSVACDAILV.....QICQEVANWDGFDVLHAPC 545

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL19.*
1: sp-archaea.*
2: sp-bacteria.*
3: sp-fungi.*
4: sp-human.*
5: sp-invertebrate.*
6: sp-mammal.*
7: sp-mhc.*
8: sp-organelle.*
9: sp-phage.*
10: sp-plant.*
11: sp-rodent.*
12: sp-virus.*
13: sp-vertebrate.*
14: sp-unclassified.*
15: sp-rviro.*
16: sp-bacteriap.*
17: sp-archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------|--------------------|
| 1 | 969.5 | 33.4 | 550 | Q937L5 | Q937L5 Comamonas t |
| 2 | 895.5 | 30.9 | 549 | O50641 | O50641 rhodococcus |
| 3 | 874 | 30.1 | 561 | Q93J19 | Q93J19 rhodococcus |
| 4 | 838 | 28.9 | 556 | Q9A7C5 | Q9A7C5 caulobacter |
| 5 | 833.5 | 28.7 | 529 | Q9FD13 | Q9FD13 brevibacter |
| 6 | 806 | 27.8 | 543 | Q9R2F5 | Q9R2F5 acinetobact |
| 7 | 804 | 27.7 | 543 | Q9R2F4 | Q9R2F4 acinetobact |
| 8 | 796 | 27.4 | 553 | Q9FD14 | Q9FD14 brevibacter |
| 9 | 786 | 27.1 | 543 | Q9A7A7 | Q9A7A7 caulobacter |
| 10 | 550.5 | 19.0 | 491 | Q9I218 | Q9I218 pseudomonas |
| 11 | 515 | 17.7 | 519 | Q9RKH5 | Q9RKH5 streptomyc |
| 12 | 509.5 | 17.6 | 492 | P71662 | P71662 mycobacteri |
| 13 | 497.5 | 17.1 | 524 | O53294 | O53294 mycobacteri |
| 14 | 486.5 | 16.8 | 437 | O87636 | O87636 pseudomonas |
| 15 | 481.5 | 16.6 | 603 | Q938F6 | Q938F6 rhodococcus |
| 16 | 466.5 | 16.1 | 527 | Q9I3H5 | Q9I3H5 pseudomonas |

| | | | | | | |
|----|-------|------|-----|----|--------|--------------------|
| 17 | 462 | 15.9 | 458 | 2 | O86908 | O86908 sphingomona |
| 18 | 435.5 | 15.0 | 603 | 2 | Q9RL17 | Q9RL17 streptomyc |
| 19 | 433 | 14.9 | 481 | 3 | Q9PAF1 | Q9PAF1 aspergillus |
| 20 | 428 | 14.7 | 640 | 2 | Q93TJ5 | Q93TJ5 pseudomonas |
| 21 | 362.5 | 12.5 | 489 | 16 | P96223 | P96223 mycobacteri |
| 22 | 359 | 12.4 | 499 | 16 | O9A8K5 | O9A8K5 caulobacter |
| 23 | 357 | 12.3 | 242 | 2 | Q93RL3 | Q93RL3 acinetobact |
| 24 | 342.5 | 11.8 | 499 | 16 | Q9I4D0 | Q9I4D0 pseudomonas |
| 25 | 339 | 11.7 | 833 | 16 | Q9R2T0 | Q9R2T0 deinococcus |
| 26 | 337 | 11.6 | 498 | 16 | O9A588 | O9A588 caulobacter |
| 27 | 333 | 11.5 | 508 | 2 | Q9APW3 | Q9APW3 pseudomonas |
| 28 | 326 | 11.2 | 486 | 16 | O53762 | O53762 mycobacteri |
| 29 | 297 | 10.2 | 495 | 16 | O53300 | O53300 mycobacteri |
| 30 | 249 | 8.6 | 406 | 10 | O9ASA7 | O9ASA7 oryza sativ |
| 31 | 240.5 | 8.3 | 424 | 10 | O9LKC0 | O9LKC0 arabidopsi |
| 32 | 238.5 | 8.2 | 354 | 2 | O9LIX3 | O9LIX3 streptomyc |
| 33 | 235.5 | 8.1 | 327 | 2 | O93JC8 | O93JC8 streptomyc |
| 34 | 232 | 8.0 | 431 | 10 | O49312 | O49312 arabidopsi |
| 35 | 225 | 7.8 | 453 | 10 | O9FKE7 | O9FKE7 arabidopsi |
| 36 | 219 | 7.5 | 458 | 16 | O9AA34 | O9AA34 caulobacter |
| 37 | 218 | 7.5 | 535 | 11 | O9Q2F7 | O9Q2F7 mus musculu |
| 38 | 216.5 | 7.5 | 414 | 10 | O9SZY8 | O9SZY8 arabidopsi |
| 39 | 216.5 | 7.5 | 532 | 6 | O95LA1 | O95LA1 canis fami |
| 40 | 216 | 7.4 | 530 | 10 | O9LMA1 | O9LMA1 arabidopsi |
| 41 | 212 | 7.3 | 426 | 10 | O9SVU0 | O9SVU0 arabidopsi |
| 42 | 212 | 7.3 | 528 | 5 | Q20730 | Q20730 caenorhabdi |
| 43 | 210.5 | 7.3 | 471 | 4 | O9BRX1 | O9BRX1 homo sapien |
| 44 | 210.5 | 7.3 | 479 | 16 | O98E17 | O98E17 rhizobium l |
| 45 | 210 | 7.2 | 421 | 10 | O64489 | O64489 arabidopsi |

ALIGNMENTS

RESULT 1

ID Q937L5 PRELIMINARY: PRT; 550 AA.
AC Q937L5; DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CYCLOPENTANONE 1,2-MONOXYGENASE (EC 1.14.13.22).
GN CPMA.
OS Comamonas testosteroni (Pseudomonas testosteroni).
OC Bacteria; Proteobacteria; beta subdivision; Comamonadaceae; Comamonas.
OX NCBI_TaxID=285;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 9872;
RA van Bellen J.B., Fritsche U., Seeger M., Smits T.H.M., Witholt B.;
RT "Cloning of Baeyer-villiger monooxygenases from Comamonas,
RT Xanthobacter and Rhodococcus via PCR with highly degenerate primers.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ418060; CADI0798.1; -
KW Monooxygenase; Oxidoreductase.
SQ SEQUENCE 550 AA; 62111 MW; 89566F95203F6173 CRC64;

Query Match 33.4% Score 969.5; DB 2; Length 550;
Best Local Similarity 36.3%; Pred. No. 1e-71;
Matches 189; Conservative 99; Mismatches 228; Indels 5; Gaps 3;
QY 5 SNSRGDCSVACDAIIVGAGLSGISAIVYKRLRLINAKIFEGAPDFGVVHNNRYPGARVD 64
DB 15 NNSVND---KLDVLLIGAGFTGLYVHLRKLGYKVLVDAGADIGGIWHNWCYPGARVD 71
QY 65 SETPPYQLNPEVKNWDCVRYDPDKELSYVHCKIRGLKRVYFGAEVVDARYAD 124
DB 72 THICQIYQISIFELQEFNWKELFPNWAQMRREYFADKLDLSKDISFNTRVQSAVFDEG 131
QY 125 LGTWTWVTSAGHATATKYLIATLGLLHRKITPALPLGLADFNKGVIHSSAW-HEDFDAEQ 183
DB 132 TREWTVRGIGHQIOAREVIANLGFAGSPSTPNVDGIEFKGWYHTALMPQEGVNMAGK 191

QY 184 RVAVTGAGATSIQIIVQELAKKADQVTMEMRPSYCLPMRQRTMDRNEQYAKKAYPTLFE 243
 DB 192 RVAVTGSSGVQYAAALDAKQVYQVTPMLALPMHOKOLSEADNLRKPELPAAFE 251
 QY 244 ASRSRGTFVQAPSVGIFVSPQREAYFEELWEGAFNFIACQYREVMDKKANRLVY 303
 DB 252 RRGCEAGFDFTAKNAKNTSAAERTTEILEELNAGGFRYLANFDFDKANDVYV 311
 QY 304 DFMAKTRRSRVNPAKRLMAPLEPPYWFCTKRSPLESYDYEMLDKPSVEIVNLQSPIV 363
 DB 312 EFWRDYVARIKDPKVAELAPKKPKPYCAKPSLEQWYEIFQNNVTLVDVNETPVL 371
 QY 364 AVTITGVLLSDGSKREDTIVLATGDSFTGSLTHGLKANKGVLDKEWKDGIYSTYMGV 423
 DB 372 RITEKGIYATG-BAEFDLIVATGDAVYTGGLTSDFRNNGQSGFKVNSDGIQTGV 430
 QY 424 FSHGFPNFAFFVATAQAPTIVLSNGPTIETQVDLIADTIKLEAEHATSVEATKSAQEAWS 483
 DB 431 ATAGFPNLLFGYQSPAGFCNGPSSAEYQGDLLIQLMNYLRDNNISRIEAQSAQEAWS 490
 QY 484 IMIAKMEHTLFLPLTDSMTTGGNIPGKATPALTFFIGGIALY 524
 DB 491 KLIADFWSLFPRAKMSYOGNIPGKVESLNFPLGLPTY 531

RESULT 2
 ID O50641 PRELIMINARY; PRT; 549 AA.
 AC O50641;
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE STERIOD MONOOXYGENASE.
 OS Rhodococcus rhodochrous.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
 OX NCBI_TaxID=1829;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IFO3338;
 RA MEDLINE=99398443; PubMed=10467180;
 RX Morii S., Sawamoto S., Yamauchi Y., Miyamoto M., Iwami M., Itagaki E.;
 RT "Steroid monooxygenase of Rhodococcus rhodochrous: Sequencing of the
 genomic DNA, and hyperexpression, purification, and characterization
 of the recombinant enzyme."
 RL J. Biochem. 126:624-631(1999).
 DR EMBL; AB010439; BAA24454.1; -.
 DR InterPro; IPR000960; FMO.
 DR Pfam; PF00743; FMO-like; 1.
 KW Monooxygenase.
 SQ SEQUENCE 549 AA; 60136 MW; 232D90D52D8E6D01 CRC64;

Query Match 30.9%; Score 895.5; DB 2; Length 549;
 Best Local Similarity 36.3%; Pred. No. 1.4e-65;
 Matches 192; Conservative 103; Mismatches 227; Indels 7; Gaps 6;

QY 16 DAIVGAGLSGISAVYKRLRLNAKIFEGAPDFGCVHWNRYPCARVDSETPPYQLNI- 74
 DB 23 DVVVVAGIAGLYAIHRISQGLTVRAFEAASGVGVWYNNRYPCARCDVESIDYSYFS 82
 QY 75 PEVMDKWTWCRCRYPOKELLSYVHCDKIRGLKDVYFGAEVVDARYARDLGTWYKTS 134
 DB 83 PELQEWNNSKYATQPEILAYIEHVADRFDLRRDITRTVTSVILDEGLRWVTRDR 142
 QY 135 GHVATKYLIATGLLHRKHTPALPGLADPNKGKVIHSSAW-HEDFAEQRVAVIGAGAT 193
 DB 143 GDEVSAREFLVYAGPLSNANTPAFDGLDRFTGDIHTARWPHDGVDFTKRGVGTGSS 202
 QY 194 STQIVQELAKKADQVTMEMRPSYCLPMRQRTMDRNEQYAKKAYPTLFEASRKRIGFP 253
 DB 203 GQISPIIAEQAEQLFFVQFSANYSIPAGNVPLDDATRAEQKANYAERRRLSREGSGSP 262

QY 254 VOAPSVGIFEVSPQREAYFEELWEGAFNFIACQYREVMDKKANRLVYDFWAKKTRSR 313
 DB 263 HRPKPSALENSEERRAVYERKWLGVLP-SKAFDQLTDPANNDTARAFWEKIRAV 321
 QY 314 INVPAKRLMAPLEPPYWFCTKRSPLESYDYEMLDKPSVEIVNLQSPIVAVTKTGVLLS 373
 DB 322 VDDPAVAELLTPKD-HAIGAKRIVTDSGYEYTYNRRONVELDURSTPIVGMDETGI-VT 378
 QY 374 DGSKRECDTIVLATGDSFTGSLTHGLKANKGVLDKEWKDGIYSTYMGVFSFGFPNAPF 433
 DB 379 TGAHYLDLWIVLATGDAWTSGLDKLEIVGRCGRTLKETWAAAGPRTYGLGIDGFPNFFN 438
 QY 434 VATAQAPTIVLSNGPTIETQVDLIADTIKLEAEHATSVEATKSAQEAWSIMAKMNEHT 493
 DB 439 LTGPGSPSLANWYHSELHVDWADATAYLDARGAAGIEGTPPAVDADWVEECNRRAEAS 498
 QY 494 LPLTDSMTTGGNIPGKATPALTFFIGGIALYEQICEKV-ANNDGFPVL 541
 DB 499 LLNSANSWYLGANIPGRPRVPMPLFGGFGVYREIITEVAESGYKGFAL 547

RESULT 3
 ID Q93JR9 PRELIMINARY; PRT; 561 AA.
 AC Q93JR9;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE BAeyer-VILLIGER MONOOXYGENASE HOMOLOGUE.
 OS Rhodococcus erythropolis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
 OX NCBI_TaxID=1833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DC114;
 RA Van der Vlugt-Bergmans C.J.B., Leer R.J., Gottschalk G.,
 RX van der Werf M.J.;
 RT "The genes involved in limonene and carveol degradation in Rhodococcus
 erythropolis DC114 are localized on a large linear plasmid."
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ303350; CAC40956.1; -.
 KW Monooxygenase.
 SQ SEQUENCE 561 AA; 62629 MW; D1D893C77F4932A1 CRC64;

Query Match 30.1%; Score 874; DB 2; Length 561;
 Best Local Similarity 34.4%; Pred. No. 8.8e-64;
 Matches 183; Conservative 101; Mismatches 228; Indels 20; Gaps 8;

QY 16 DAIVGAGLSGISAVYKRLRLNAKIFEGAPDFGCVHWNRYPCARVDSETPPYQLNI- 74
 DB 15 DVLVVAGSGGRLALHLREKCYRTHLIEAGDDVCGCTWYNNRYPCARVDLESYAFD 74
 QY 75 PEVMDKWTWCRCRYPOKELLSYVHCDKIRGLKDVYFGAEVVDARYARDLGTWYKTS 134
 DB 75 DELQLEWEMPERYAAQSDVRLYLQHVADRFDLRRDISFGERVQSATYDEASATWLVTDC 134
 QY 135 GHVATKYLIATGLLHRKHTPALPGLADPNKGKVIHSSAMHE-DFDAEQRVAVIGAGAT 193
 DB 135 GEQWRSRPFIPAVGFLSTPYVPDIQGRESFAGQLVHSARWPEGGIDFACKRAVITGTST 194
 QY 194 STQIVQELAKKADQVTMEMRPSYCLPMRQRTMDRNEQYAKKAYPTLFEASRKS----- 248
 DB 195 GVOLIPIAESASLTVLQSPMMVVPLOQNTPMPEEYQTRIKARYP---ELRRELDSEF 251
 QY 249 ----RIGFPVQAPSV-GIFEVSPEQREAYFEELWEGAFNFIACQYREVMDKKANRLVY 303
 DB 252 AGNLLVDFDLRSSETRSALVSPEREAEYDFRWAGGLS-LYTSFTDLIFNQEANDTLR 310
 QY 304 DFMAKTRRSRVNPAKRLMAPLE-PPYWFCTKRSPLESYDYEMLDKPSVEIVNLQSP 362

```

Db 311 EYLERRIRQLIKDPATADKLIPTDYPPL---TKRLCCERGYYDAFNRDNVHLVDTKADPI 367
QY 363 VAVTKTGVLLSDGSKRCEDTIVLATGDSFTGSLTHMGLKNKHGVDLKEVWKDGISTYMG 422
Db 368 AEITPTGVRLSGAVHDVLIIFATGDAATGSLKRLNITGPGORTLDEHWADGARTHLG 427
QY 423 VFSHGFMNAFFVAQAQATVLSNGPTIETQVDLIADTTAKLEAEHATSVEATKSNQAEAW 482
Db 428 LMSDAFPNMFPLAGPLSPGGFFSPPOSDYOVLIAARMIDRLAGEDAVAIEPEPAVEDW 487
QY 483 SIMIAKMEHTLPLTDSWMTGGNIPGKATRALTFIGGIALYBQICQEVAN 534
Db 488 MTOVDVAVNATLLPKAQSWSGANIPGKARQFLYLLGGFAAYRDAEQMFAD 539

RESULT 4
Q9A7C5 PRELIMINARY; PRT; 556 AA.
AC Q9A7C5;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE STEROID MONOOXYGENASE.
GN CC1799.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Deloy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -1- COFACTOR: FAD (BY SIMILARITY).
DR TIGR; CC1799; -.
DR InterPro; IPR001327; FAD_pyr_redox.
DR InterPro; IPR000960; FMO.
DR InterPro; IPR001100; pyr_redox.
DR Pfam; PF00743; FMO-like; 1.
DR PRINTS; PR00368; FADPDR.
DR PRINTS; PR00411; PNDROTASEI.
KW Complete proteome; FAD; Flavoprotein; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 556 AA; 60899 MW; A91CE30779702482 CRC64;

Query Match 28.9%; Score 838; DB 16; Length 556;
Best Local Similarity 33.4%; Pred. No. 8.3e-61;
Matches 179; Conservative 104; Mismatches 235; Indels 18; Gaps 6;

QY 16 DAIIIVAGLSGISAVYKLRKLRLNAKIFEGAPDEGGVWNNRYPGARVDSETPFYQLNI- 74
Db 17 DALVITGAGFGGYAVHKLSSGMLSGVGGVGGVWNNRYPGARCDLMSLDICYSPS 76
QY 75 PEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDYVFGAIVVDARYDLGTWTVTISA 134
Db 77 PDIEQEWMTSEQFAAQSEILAYANFVADRLDLRRHFLFTWTGTAAYDEANWVRVDD 136
QY 135 GHVATAKYLILATGLLHRKHHTPALPLGLADFNKGKVIHSSAW- HEDFDEAGORVAVICAGAT 193
Db 137 GRVFATTCVNASGSLSPKGVPTDGDADDFKGEILLAAKRNPHPTFVSEGRKRVGLVGTST 196
QY 194 SIQIVQELAKADQVTFMRRPSPCYCLPMQRQTMDRNPDOTAKWAKYIYPTLFASRKSRI 253
Db 197 GIQIVPEVAKTAGALVFQRTPTSLPMRNVTLPEPDVVAELKRNKYGIRQVARNSP 256

```

```

QY 254 VOAPSVGIFEVSEPOREAYFEELWEGAFNFLACQYREVWMDKKANRLLYDFWMAKTKSR 313
Db 257 RQOSTRPFPSVTPEQRALMEDSWARGLSFLG-TFSOLLVNPNEANEQVAEPVRSKISEV 315
QY 314 IVNPAKRDLMAPLEPPVFGTKRSPLESYVEMLDKPSVEIVNLBQSPVAVTKTGVLIS 373
Db 316 VTDPVTAELKPGYPIF--ARRPCLDTHYETFNLPVHLKLDCLSPVKITPAG-LLT 372
QY 374 DGSKRECDTIVLATGDSFTGSLTHMGLKNKHGVDLKEVWKDGISTYMGVFSHGFPNAPF 433
Db 373 ETGEVELDLVIFATGDLTGALLNFDVVGRLQURDKWDGAKSHLLGIMTASFPNLF 432
QY 434 VATAQAPTVLSNGPTIETQVDLIADTTAKLEAEHATSVEATKSAQEAWSIMIAKMEHT 493
Db 433 VCGPNGPAALANIITLDOQNIDWADAIHMRQEGLATIETPEGAQGGWMDVAVYMAELT 492
QY 494 LFLTDSWMTGGNIPGKATRALTFIGGIALYEQICQ-----EKVANWDG 537
Db 493 LVSKANTWYVGNISGNPKRGLSMYTGFGQRYSEACRLAAAEKGYEDFVFKEAATKQG 548

RESULT 5
Q9FDI3 PRELIMINARY; PRT; 529 AA.
AC Q9FDI3;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CYCLOHEXANONE MONOOXYGENASE 2.
GN CHNB2.
OS Brevibacterium sp. HCU.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococccineae; Brevibacteriaceae; Brevibacterium.
OX NCBI_TaxID=133406;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCU;
RX MEDLINE=20353458; PubMed=10894733;
RA Brzostowicz P.C., Gibson K.L., Thomas S.M., Blasko M.S.,
RA Rouviere P.E.;
RT "Simultaneous identification of two cyclohexanone oxidation genes from
RT an environmental Brevibacterium isolate using mRNA differential
RT display.";
RT J. Bacteriol. 182:4241-4248(2000).
CC -1- COFACTOR: FAD (BY SIMILARITY).
DR EMBL; AF257215; AAG01290.1; -.
DR InterPro; IPR001327; FAD_pyr_redox.
DR InterPro; IPR001100; pyr_redox.
DR PRINTS; PR00368; FADPDR.
DR PRINTS; PR00411; PNDROTASEI.
KW FAD; Flavoprotein; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 529 AA; 59143 MW; 71DE09C84E441BF CRC64;

Query Match 28.7%; Score 833.5; DB 2; Length 529;
Best Local Similarity 32.5%; Pred. No. 1.8e-60;
Matches 167; Conservative 107; Mismatches 237; Indels 3; Gaps 3;

QY 5 SNSRGDCSVACDAIIIVAGLSGISAVYKLRKLRLNAKIFEGAPDEGGVWNNRYPGARVD 64
Db 11 AQANADETEYDLALIVGGFSGVPSVDRLRDEGDFKVKYVMDAAGGFGGIIWNNCYPGARTD 70
QY 65 SETPFYQLNTPVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDYVFGAIVVDARYD 124
Db 71 STGOIYQFOYKDLKKDFEKLFPDFNGVREYFVDSQLDSRDVTENTFAESCTWDDA 130
QY 125 LGTWTVTISAGHVATKYLILATGLLHRKHHTPALPLGLADFNKGKVIHSSAWHE-DFDAEQ 183
Db 131 AKETVTSRSEGREQARAVATVFGAKPLYPNIEGLDSFEGECHHTARWPOGGLDWTGK 190
QY 184 RVAVIGAGATSIQIVQELAKADQVTFMRRPSPCYCLPMQRQTMDRNEQATWAKYIYPTLFE 243

```

DB 191 RVVNGTGAAGTQVTOEAAVAEHLTVQPTNLAALPMROORLSADDNDRIYRENIERFQ 250
QY 244 ASRKSRIIGFPVQAPSGVIGFVSPQREAYFPEELWEGAFNPLACQYREVVMVDKKANRLVY 303
DB 251 IRDSNFAAGDFVFTFQNAADTPEDERTAIYEKMDGEGFPLWLGNFQGLLTDEAANTFY 310
QY 304 DEWAKTRSRIVNPAKRDLMAPLEPPYFCTGKRSPLSDIYEMLDKPSVEIVNLQSPIV 363
DB 311 NFWRSKVDHVRDKPTAKDMLAPATPPHFGVGRPSLSONYFDVYNQDNVDLIDSNATPIT 370
QY 364 AVTKGVLLSDGSKRECDTIVLATGDSFTGSLTHMGLKKNKHGVDLKEVWKGDIYSTYMGV 423
DB 371 RVLPMGVETPGVW-ECVDVLVATGFDNNSGGINAIDIK-AGGQLLRDKWATGVDTYMG 428
QY 424 FSHGPPNAFFVATQAQPTVLSNGPTIETQVLDLADITAKLEAEHATSVETKSAQEAWS 483
DB 429 STHGPPNLFYLGQSPSGFCNGTDFGAPGDMVADFLWLKDNIGRSFSTEEVEWR 488
QY 484 IMAKMHHTLPLTDSWMTGNGIPGNATRALTF 517
DB 489 AHVDDIFVNSLFPRAKSWYGANVPKPAOMLNY 522

RESULT 6
Q9R2F5 PRELIMINARY; PRT; 543 AA.
AC Q9R2F5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CYCLOHEXANONE 1,2-MONOXYGENASE.
GN CHNB.
OS Acinetobacter sp., and
OC Acinetobacter sp., NCIMB9871.
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
OC Acinetobacter.
OX NCBI_TaxID=472, 93373;
[1]
RN SEQUENCE FROM N.A.
RP SPECIES=Acinetobacter sp.; STRAIN=NCIB 9871;
RX MEDLINE=20011264; PubMed=10543838;
RA Iwaki H., Hasegawa Y., Teraoka M., Tokuyama T., Bergeron H.,
RA Lau P.C.K.;
RT "Identification of a transcriptional activator (Chnr) and a 6-
RT oxohexanoate dehydrogenase (Chne) in the cyclohexanol catabolic
RT pathway in Acinetobacter sp. strain NCIMB 9871 and localization of the
RT genes that encode them.";
RL Appl. Environ. Microbiol. 65:5158-5162(1999).
[2]
RN SEQUENCE FROM N.A.
RP SPECIES=Acinetobacter sp.; STRAIN=NCIB 9871;
RX MEDLINE=88115180; PubMed=3338974;
RA Chen Y.-C.J., Peoples O.P., Walsh C.T.;
RT "Acinetobacter" cyclohexanone monooxygenase: Gene cloning and sequence
RT determination.";
RL J. Bacteriol. 170:781-789(1988).
[3]
RN SEQUENCE FROM N.A.
RP SPECIES=Acinetobacter sp. NCIMB9871; STRAIN=NCIMB 9871;
RA Iwaki H., Hasegawa Y., Teraoka M., Tokuyama T., Lau P.C.;
RT "Identification and Characterization of Cyclohexanol Metabolic Genes
RT from Acinetobacter sp. NCIMB 9871.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- COFACTOR: FAD (BY SIMILARITY).
DR EMBL; AB006902; BAB86293.1; -;
DR EMBL; AB026668; BAB61738.1; -;
DR InterPro; IPR001327; FAD_pyr_redox.
DR InterPro; IPR000960; FMO.
DR InterPro; IPR001100; pyr_redox.
DR Pfam; PF00743; FMO-like; 1.
DR PRINTS; PR00368; FADPNR.
DR PRINTS; PR00370; FMOXYGENASE.
DR PRINTS; PR00411; PNDRDTASE1.

KW FAD; Flavoprotein; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 543 AA; 60937 MW; EC9A2ABE49A270AF CRC64;

Query Match 27.8%; Score 806; DB 2; Length 543;
Best Local Similarity 33.2%; Pred. No. 3.6e-58;
Matches 179; Conservative 100; Mismatches 230; Indels 30; Gaps 8;

QY 16 DAIIAGAGLSGISAVYKLR-KLRNNAKIFEGAPDFGGMVHNNRYPGARVDSETPFY---- 70
DB 8 DAIVGGGGLAVAKRLDELKLVQAFDKATDVAGTWNRYNRYPGALDTDETHLYCYSW 67
QY 71 --QLNIPVWKDWTSCRYPDOKELLSYVHCDKIRGLRKDVYFCAEYVDARYADLGT 127
DB 68 DKELLOSLEIKK-----KYVQGPVRYKLYQOVAEHLKSKYQNTAVQSAHYNEADAL 121
QY 128 WTVKTSAGHVATAKYLILATGLLHRRHTPALPGLADFNKVIHSSAWHEDFDAGORVAV 187
DB 122 WEYTYEGDKYTARFLITALLGLLSAPNLPNIKINGINOFKGLHHTSRMPDDVSFEGRGV 181
QY 188 IGAGATSIQVQELAKKAQVTFMRRPSCYCLPMQRTMDRNEQTAKKAYIPTLFEASRK 247
DB 182 IGTGSGVQVITAVAPLAKHLTVFQSAQSYVPIGNPLSEEDVKKIKONYDKIMGVWN 241
QY 248 SRIGFPVQAPSGVIGFVSPQREAYFPEELWEGAFNPLACQYREVVMVDKKANRLVYDFW 306
DB 242 SALAFGLNESTVPAMSVASPERKAVFEKAWGTGGGFRFMFTFGDIATNNEANIEAQNFI 301
QY 307 AKTRSRIVNPAKRDLMAPLEPPYFCTGKRSPLSDIYEMLDKPSVEIVNLQSPIVAVT 366
DB 302 KGKIAETVDPALFAQKLMPOD----LYAKRPLCDSGYNTFNDRNDVRLSDVKNPIVEIT 357
QY 367 KTCVLLSDGSKRECDTIVLATGDSFTGSLTHMGLKKNKHGVDLKEVWKGDIYSTYMGVFSH 426
DB 358 ENGVRLNGDGFVELDMLICATGDAVDGNYVRMDIOCKNGLAMKDYMKGSPSYMGVTYN 417
QY 427 GFPAEFVATAQAAPTIVLSNGPTIETQVLDLADITAKLEAEHATSVETKSAQEAWSIMI 486
DB 418 NYPNMEMVLGPNGP--FTNLPESIESQVEWISDTIOTVYNNVNESTEATKEAEQWOTQC 475
QY 487 AKNEHTLPLTDSWMTGNGIPGNATRALTFITGGIALYEQICQEKVAN----WDGFV 540
DB 476 ANIAENTLFPKAOSWIFGANIPGKKNTVYFYLGLKEY----RSALANCKNHAYEGFDI 530

RESULT 7
Q9F7E4
ID Q9F7E4 PRELIMINARY; PRT; 543 AA.
AC Q9F7E4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CYCLOHEXANONE MONOOXYGENASE.
GN CHNB.
OS Acinetobacter sp. SE19.
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
OC Acinetobacter.
OX NCBI_TaxID=135835;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=SE19.
RX MEDLINE=20398154; PubMed=10940013;
RA Cheng Q., Thomas S.M., Kostichka K., Valentine J.R., Nagarajan V.;
RT "Genetic analysis of a gene cluster for cyclohexanol oxidation in
RT acinetobacter sp. strain SE19 by in vitro transposition.";
RL J. Bacteriol. 182:4744-4751(2000).
CC -1- COFACTOR: FAD (BY SIMILARITY).
DR EMBL; AF282240; AAG10021.1; -;
DR InterPro; IPR001327; FAD_pyr_redox.
DR InterPro; IPR000960; FMO.
DR InterPro; IPR001100; pyr_redox.
DR PRINTS; PR00368; FADPNR.
DR PRINTS; PR00370; FMOXYGENASE.

DR PRINTS; PR00411; PNDRTASEI.
KW FAD; Flavoprotein; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 543 AA; 60923 MW; 77E1EC0E22A4C0C5 CRC64;

Query Match 27.7%; Score 804; DB 2; Length 543;
Best Local Similarity 33.2%; Pred. No. 5,3e-58;
Matches 179; Conservative 100; Mismatches 230; Indels 30; Gaps 8;

QY 16 DAIVGAGLSGSAVYKLR-KLRNAKIFEGAPDFGCVWNNRYPGARVDSETPFY----- 70
DB 8 DAIVGGFGGLYAVKLRDELEKVAQFADKATDVAGTWYNNRYPGALSDDTETHLYCSW 67
QY 71 ---OLNPEYKWTSCRYPDQKELLSYVHHCCKIRGLKRDVYFGAEDVARYARDLGT 127
DB 68 DKELLQSLKIKK-----KYVQGPDKVKYLOQVAEKHDLKKSYPQNTAVQSAHNEADAL 121
QY 128 WTKVTSAGHVATAKYLILATGLLHRKHTPALPLGLADFNKGKVIHSSAHWEDFAEGORVAV 187
DB 122 WEVTEYGDYKATFALITAGLLSAPNLNIGINQFNGELHHSRWPDVDSFEGKRVGV 181
QY 188 IGAGATSIQVOELAKKADQVTMFMRRPSYCLPMKQRTMDRNEQTAHAYPTLFEASRK 247
DB 192 IGTGSTGVVITAVAPLAKHLTVFQSAQYSPVIGNDPLSEEDVKKIKDNDKIDWDGVMN 241
QY 248 SRIGFPVQAPSVGIFESPQREAYFEELWER-GAFNFLACQYREVYVDDKRNRLVDFW 306
DB 242 SALAFGLNESTVPAMSVSAERKAVPEKANQVTGGGFRFMETFGDIATNMEANTEAQNFI 301
QY 307 AKRTFSRVNPAKEDLMAPLEPPYWGFKRSPLSDYEMLDKPSVEIVNLEQSPIYAVT 366
DB 302 KGKIAEIVKQATIAQLMPQD----LYAKRPLCDGSGYNTFNRDNRLEDVKAHPIVEIT 357
QY 367 KTGVLSDGSKRECDTIVLATGDSFTSLTHMGLKNKHGVLDKEVWKDGISTYNGVFSH 426
DB 358 ENGKLENGDEVELDMICATGDAVDGNYVRMDIQGNGLAMKMDYKREGSPSYMGVTN 417
QY 427 GFPAFFVATAQAPTIVLSNGPTTIETQVDLIADTIKLAEEHATSVKATSKAQAWSIMI 486
DB 418 NYPNMFVGLNCP--FTNLPPSIESOVESIDTIQYTVENNVSIEATKAEASQWQTC 475
QY 487 AKMHEHTLPLTDSWTTGNNIPKATRALTFGGTALYEQICEKVN-----WDGDDV 540
DB 476 ANIAEMTLFPKAQSNIGANIPKNTVTFYLGGLKEY----RSALANCKNHAYEGFDI 530

RESULT 8
ID Q9FDI4 PRELIMINARY; PRT; 553 AA.
AC Q9FDI4;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CYCLOHEXANONE MONOOXYGENASE 1 (EC 1.14.13.22).
GN CHNB1.
OS Brevibacterium sp. HCU.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococcales; Brevibacteriaceae; Brevibacterium.
OX NCBI_TaxID=133406;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCU;
RX MEDLINE=20353458; PubMed=10894733;
RA Brzostowicz P.C., Gibson K.L., Thomas S.M., Blasko M.S.,
RA Rouviere P.E.;
RT "Simultaneous identification of two cyclohexanone oxidation genes from
RT an environmental Brevibacterium isolate using mRNA differential
RT display.";
RL J. Bacteriol. 182:4241-4248(2000).
CC -!- COFACTOR: FAD (BY SIMILARITY).
DR EMBL; AF257214; AAG01289.1; -
DR InterPro; IPR001327; FAD_pyr_redox.
DR InterPro; IPR000103; Pyridine_redox_2.

DR InterPro; IPR001100; pyr_redox.
DR PRINTS; PR00368; FADPNR.
DR PRINTS; PR00411; PNDRTASEI.
DR PRINTS; PR00469; PNDRTASEII.
KW FAD; Flavoprotein; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 553 AA; 62141 MW; 84747D4974A93169 CRC64;

Query Match 27.4%; Score 796; DB 2; Length 553;
Best Local Similarity 33.7%; Pred. No. 2.5e-57;
Matches 177; Conservative 105; Mismatches 224; Indels 20; Gaps 9;

QY 16 DAIVGAGLSGSAVYKLRKLRNAKIFEGAPDFGCVWNNRYPGARVDSETPFYQOLNI- 74
DB 10 DAIVGAGFGLAHLHLREIGLDTQIVEATDGIGGTWNNRYPGVRTDSEHYISFSS 69
QY 75 PEVMDWTSCRYPDQKELLSYVHHCCKIRGLKRDVYFGAEDVARYARDLGTWTKTSA 134
DB 70 KEVRDEWWTORYPDGEVCAYLNFIAADLDLAKDIOLNSRVNTARWNETEKYWDVIFED 129
QY 135 GHVATAKYLLIATGLLHRKHTPALPLGLADFNKGKVIHSSAHWED-FAEGORVAVIGAGAT 193
DB 130 GSKRFARELISAMGALSQAIFPAIDGIDFENGAKYHTAAMPADGVDTGCKKVCYLGVAS 189
QY 194 SIQIYOELAKKADQVTMFMRRPSYCLPMKQRTMDRNEQTAWKAY-----YPTLFEASRKS 249
DB 190 GIOIPELAKLAGELFVFTPTNYVVSNNDKVD---AEMQYVRDNTDEIFERASRHP 245
QY 250 TGFVQAPSVGIFESPQREAYFEELWERGAFNFLACQYREVYVDDKRNRLVDFWMAKK 309
DB 246 FGVDMEYPTDSAVESEERKRVFESKWEBSGFHANECDTDLGTSPEASELASEFIRSK 305
QY 310 TRSRIVNPAKEDLMAPLEPPYWGFKRSPLSDYEMLDKPSVEIVNLEQSDIVAVTKTG 369
DB 306 REVVKQDAPADLLCP--KSYSFNGKRVPPGHGYYETFNFWLLDARGTPILTRISSKG 363
QY 370 VLLSDGSKRECDTIVLATGDSFTSLTHMGLKNKHGVLDKEVW-KDGISTYNGVFSHG 428
DB 364 LVHGD-TEVELDAVFATGPDAMTGTITNIDIVGRDGVILRKWAODGLRTIGLVNCF 422
QY 429 PNAFFVATAQAPTIVLSNGPTTIETQVDLIADTIKLAEEHATSVKATSKAQAWSIMIAK 488
DB 423 PN-PLMSIGSPOTPSNLVPIQLGAQMQRFLKFIQERGLIEVFESSREAEIWAETIR 480
QY 489 MNEHTLF-----PLTDSWTTGNNIPKATRALTFGGTALYEQICE 530
DB 481 GAESTVMSIEGPKAGAMPIGNIPKRSREYQVYMGQGVYQDMCRE 526

RESULT 9
ID Q9A7A7 PRELIMINARY; PRT; 543 AA.
AC Q9A7A7;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE STEROID MONOOXYGENASE.
GN CC1818.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Bodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uytterback T., Tran K., Wolf A., Vanatavan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;

01-DEC-2001 (Tremblrel. 19, Last annotation update)
DB MONOOXYGENASE.
GN SCE87.23C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL MOL. Microbiol. 21:77-96(1996).
CC -1- COFACTOR: FAD (BY SIMILARITY).
DR EMBL; AL132674; CAB59668.1; -;
DR InterPro; IPR000759; Adnrx_reductase.
DR InterPro; IPR001327; FAD_pyr_redox.
DR InterPro; IPR000960; FMO.
DR InterPro; IPR002025; NAD_binding.
DR InterPro; IPR001003; Pyridine_redox_2.
DR InterPro; IPR001100; Pyr_redox.
DR Pfam; PF00743; FMO-like; 1.
DR PRINTS; PR00419; ADXRDTASE.
DR PRINTS; PR00368; FADPR.
DR PRINTS; PR00370; FMOXYGENASE.
DR PRINTS; PR00411; PNDROTASEI.
DR PRINTS; PR00469; PNDROTASEII.
KW FAD; Flavoprotein; Oxidoreductase.
SQ SEQUENCE 519 AA; 56673 MW; 4031FB54427A57B4 CRC64;

Query Match 17.7%; Score 515; DB 2; Length 519;
Best Local Similarity 29.5%; Pred. No. 4.4e-34;
Matches 151; Conservative 75; Mismatches 189; Indels 96; Gaps 18;

QY 19 IVGAGLSGISAVYKLRLLNAKIFEGAPDF-----GGVHWNRYPGARVDSPTPF 69
DB *15 VIGSGFGIGLNAVRLRR-----EGTDFVVLERAGSVGTWRDNSYFGACDVPSHL 66

QY 70 YQINL---PEVWKNWTSYRPOKELLXVHCHDKIRGLKDVYFGAEVVDARYDLG 126
DB 67 YSFAPNPE-----WPRFSGQHIRAYLEHVADTFGLRPLHFRDSEVKRMANDTEQL 120

QY 127 TWVTKTSGHVATAKYILATGLHRRKHTPALPLADFNKGKVLTHSSAWHEDFDAQRYA 186
DB 121 RWEIETVRG-TLTADVVSATGPLSDPKVPDIFGLDTFPGKVFHSARWDHVDLAGQRYA 179

QY 187 VIGAGATSIQIYQELAKADQVTFMRRRSYCLPMQRQMTWRDNEQTAKYVPTLFEASR 246
DB 180 MIGTGAISQIYVSIQPKVRLTLFQTPAWMPRVDRYDAISGAERALHRA-----LPATT 234

QY 247 KSRIGFPVQAPSVGIFSVSPQREAYFEELWEGAFNFLACQYREVMVD---KXANRLVY 303
DB 235 KLRRGL-----LW-----GTRELQVQAFKHPNELGF 261

QY 304 DFWAKKTSRIVNPAKRD--LMALEPPYWFQTKRSPLESYDEYMLDKPSVIVNLEOSP 361
DB 262 ---VEQIAKRNNGAAIPALRAKLTPDYKIGCKRILLSTSTYPALAKPNVDV---ASG 315

QY 362 IVAVTKTGVLSDGSKRECDTIVLATGFSFTGSLTHMGLKNNK---HGVDLKEYWKDGI 417

DB 316 LSEVRGSLVAADGTAEADAIVFCTGF-----HVTDMPIAERVVGADGRTLAETWKGGM 370
QY 418 STYMGVESHGFPNAPFVATAQAFTVLSNGPTI--TETQVDLIADTIKLEA-EHATSVEA 474
DB 371 EALRGDTAGFPN--FMTVIGPNTGLGSSMILMIESQLNYLADLYRLNVLGGRTALDP 428
QY 475 TKSQAQAWSIMIAKMEHTLFLPLTDSWMTGG 505
DB 429 RPAARVNNHRYOERMKRTV-----WNTGG 453

RESULT 12
ID P71662 PRELIMINARY; PRT: 492 AA.
AC P71662;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PUTATIVE MONOOXYGENASE CY21B4.10C (EC 1.14.13.-).
GN RV1393C OR MTCY21B4.10C
OS Mycobacterium tuberculosis
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala E.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies T., Devlin K., Feltham D., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RL Nature 393:537-544(1998).
CC -1- COFACTOR: FAD FLAVOPROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: TO M.TUBERCULOSIS MTCY31.20 AND EMERICELLA NIDULANS
CC STCW.
DR EMBL; Z80108; CAB02175.1; -;
DR Tuberculin; RV1393C; -;
DR InterPro; IPR000960; FMO.
DR Pfam; PF00743; FMO-like; 1.
KW Hypothetical protein; Oxidoreductase; Monooxygenase; NADP;
KW Flavoprotein; FAD; Complete proteome.
FT NP_BIND 171 203
FT NADP (BY SIMILARITY).
SQ SEQUENCE 492 AA; 55251 MW; 9E8FCDABEFF70EABC CRC64;

Query Match 17.6%; Score 509.5; DB 16; Length 492;
Best Local Similarity 27.8%; Pred. No. 1.2e-33;
Matches 143; Conservative 77; Mismatches 226; Indels 69; Gaps 14;

QY 17 AITVAGLSGISAVYKLRLL--NAKIFGAPDFGGVHWNRYPGARVDSPTPFVQLNIP 75
DB 7 ALIVAGFSGIGNAKILDRAGFSDYLVVAGDGVGTWHTWNTYPCJADYDPSFYQSFSE 66

QY 76 EVKNDWTSYRPOKELLXVHCHDKIRGLKDVYFGAEVVDARYDLGTWVTKTSAG 135
DB 67 Q-----SRHWSRTYAPGHELKAYABCHVCYKGIKRLNATKYLAEFDEHSLWRVQDTPG 123

QY 136 HVATKYLIATGLLHRRKHTPALPLADFNKGKVLTHSSAWHEDFDAQRYAVIGAGATSI 195
DB 124 GEITARELISACGILVAPKLPDIDGVDSFEGVTMTARWDHTQDLTKRVRGIIGTGASAV 183

QY 196 QIVQELAKADQVTFMRRRSYCLPMQRQMTWRDNEQTAKYVPTLFEASR 246
DB 184 QVPEMAPIVSHLTVFQRTPIWCFP----KFDVPLPTAVRWAMRIPGGKAVHRLLSQAFV 239

```
Qy 247 KSRIGFVQAPSQIFVSPQEQREAYFEELWERNFACQYREVMVDKKNRLYYDFW 306
Db 240 EAT--FFIAAHYAFPLAKHMSA-----G 263
Qy 307 AKTRSRIVNPAKRDLMAPLEPPYWFCTKRSPLSDYEMLDKPSVEIVNLEQSPVAVT 366
Db 264 RYLRQVHDVPYRE---QLTPRYAVGCKRPGFHWYLSLTFNRDNLVLT--EPIDKIT 317
Qy 367 KTGVLSDGSKRECDTIVLATGFDSP--TGSLSLTHMGLKNKHGVLDKEVMD--GISTYMGVF 424
Db 318 PTAVATTDGASHEDVLVLTATGFKVLDTSIPTAYVTGTGASLSRFDHHRQAEGYS 377
Qy 425 SHGPPNAFFVATAOAPTVLSNGPTIETQVDLIADITAKLEAEHATSYEATKSAQZAWSI 484
Db 378 VPGYPN--FTTVFGYGVGSSYFALITQAHHIIIRCLKRRARRGAYRIVETEARARYA 436
Qy 485 MIAKNHEHTLF-----PLTDSWN--TGGNIPGKAT 512
Db 437 EVMRHRHQVFWQDSCLANSYTFDKNGDVPURPT 471

RESULT 13
ID O53294 PRELIMINARY; PRT; 524 AA.
AC O53294;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE MONOXYGENASE.
GN RV3049C OR MT012.64C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL: AL021287; CAA16134.1;
DR TuberculList: RV3049c;
DR InterPro: IPR000960; FMO.
DR Pfam: PF00743; FMO-like; 1.
KW Complete proteome.
SQ SEQUENCE 524 AA; 58776 MW; BABAF86F4DD1AB54 CRC64;
```

```
Query Match 17.1%; Score 497.5; DB 16; Length 524;
Best Local Similarity 28.7%; Pred. No. 1.3e-32;
Matches 142; Conservative 78; Mismatches 218; Indels 57; Gaps 13;

Qy 17 AIIVAGLSGISAVYKLRKLRLNAKIFEGAPDFGCVHNNRYPGARVDSPTPPQNLNIP 76
Db 25 AVIIGTGFSLGMAIALQGVDFVILEKADDDYGGTWRDNTYPGACDIPSHLYSFSF-E 83

Qy 77 VKMDWTSCRYPOKELLSYVHCHDKIRGLKDVYFAGVVDARYARDLGTWTKTSAGH 136
Db 84 PKADWKILFSYWD--ETILGKGVTKYGLRRYIEFNSLVDRGYDDECRWHVFTADGR 141

Qy 137 VATAKYLILATGLLHRKHTPALGLADFNCKVTHSSAWHEDFAEGORVAVICAGATSIQ 196
Db 142 EYVAFVLSAGALHIPSFFIAGRDEEAGPAPFSAQWDHSDILTKRVAIVGTGASATQ 201
```

```
Qy 197 IVOELAKKADQVTFMRRPSYCLPMRQRTMDRNEQTAMKAYPTLFEASRSKRIGFPVOA 256
Db 202 IYPEIVGVQVAELQLYQRTPPWVPRTNBELPVSLLRRALT--VPLGLALLR----- 250
Qy 257 PSVGIFEVSPEQREAYFEELWERNFACQYREVMVDKKNRL--VYDFWAKTRSRIV 315
Db 251 --LGII-----NAQEALAY-----GHTKRPNTLKIIENAYAKYNIIRSV 286
Qy 316 NPAKRDLMAPLEPPYWFCTKRSPLSDYEMLDKPSVEIVNLEQSPVAVTGTGVLSDG 375
Db 287 K--DRELARKLTTPRYIGCKRILNSSTYYPVAVADPKTELIT---DRIDRITHDGIIVTADG 341
Qy 376 SK---RECDDTVLATGFDSPFTGSLTHMGLKNKHGVLDKEVMD--KDCIGTYMGVFSHGFPN 430
Db 342 TGREVFREADVIVATGF--HYTDSYTYVQIKRGHGEDLVDRNREGICAHRGITVANMPN 400
Qy 431 AFFVATAOAPTVLSNGPTIETQVDLIADITAKLEAEHATSYEATKSAQZAWSIAMIKN 490
Db 401 LFFLLGPNTGLGHSNVFWMIESQIHVADATAKCDRMGVQALAPTREADO-----RFN 453
Qy 491 EHTLFLPLTDSMMWTG 505
Db 454 QELQRRLAGSVWNSG 468

RESULT 14
ID O87636 PRELIMINARY; PRT; 437 AA.
AC O87636;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CYCLOHEXANONE MONOOXYGENASE HOMOLOG (FRAGMENT).
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=294;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=DSM50106;
RA MEDLINE=99124577; PubMed=9925571;
RA Khlamayzer V., Fischer I., Bornscheuer U.T., Altenbuchner J.;
RT "Screening, nucleotide sequence, and biochemical characterization of
RT an esterase from Pseudomonas fluorescens with high activity towards
RT lactones."
RL Appl. Environ. Microbiol. 65:477-482(1999).
DR EMBL: AF090329; AAC36351.1;
DR InterPro: IPR000960; FMO.
DR Pfam: PF00743; FMO-like; 1.
DR PRINTS: PR00469; PNDRTASEII.
KW Monooxygenase.
FT NON-TER 437
SQ SEQUENCE 437 AA; 48863 MW; 11E72DF34C0CD13C CRC64;
```

```
Query Match 16.8%; Score 486.5; DB 2; Length 437;
Best Local Similarity 27.6%; Pred. No. 7.7e-32;
Matches 137; Conservative 79; Mismatches 204; Indels 77; Gaps 14;

Qy 1 MSNTSNSRGDSVACDAIIVAGLSGISAVYKLRKLRL--NAKIFEGAPDFGCVHNNRY 59
Db 1 MNAHSDS-----IDIAIIGSGFAGLCMAIKLEAGTDLFVAFQADTLGGTWRDNLNYP 53

Qy 60 GARVDSETPPYQNLN---PEVMDWTSCRYPDQKELLSYVHCHDKIRGLKDVYFAGV 116
Db 54 GCACDVQSHVYSFSFAPNDP-----WTRQPAQAEIRAYLEDCAVREGLAPYLRFGMGL 107

Qy 117 VDARYARDLGTWTKTSAGHVATAKYLILATGLLHRKHTPALGLADFNKGVIVHSSAWHE 176
Db 108 KRAVDEQLQRQLSFGSRHVSARVLVSGMGALARPALPEIPGLTETPKGRFRFSQWDH 167

Qy 177 DFDAGQRVAVIGAGTATSIQIVELAKKADQVTFMRRPSYCLPMRQRTMDRNEQTAMKA 236
```

```

Db 168 AYALGKRRAVIGTGAIAQVQIAPOVAHLDLQRTPPWIMKPDGRGISAPEF--W-- 223
QY 237 YYTLEASRKSRIIGFVQAPSQIFVSPQREAYFEELWERAFLACQREV----- 292
Db 224 ----LFR-----HLPTQRLV-----RGAP-YWALEGRVLGAL 252
QY 293 -----MVKKANRLVYDFWAKKTRSRIVNPAKRDLMAPLEPPYKRSPLSDYYEM 346
Db 253 HPQLMKVQKVALR-----HLRKQVPRPS---LRKALTPDYTGCKRVLISNDYYP 301
QY 347 LDRPSVEIVNLEQSPIVAVTKTGVLLSDGSKRECDTIVLATGPDSTCSLTHMGLKNKHG 406
Db 302 LSRNVEVVT---DKILRIEADGVITADGKHPADCLIFGTGQA-TDPLPRDCIIGRDG 357
QY 407 VDLKEVWKDGISTYMGVFSHGFPNFAFVATAQAFTVLSNGPTIETQVDLIADTIAKLEA 466
Db 358 VDLMDTWDRGAHAKYGTIVPCYPNLFLLIGPNTGLGHNSMILMIEAQVYIILDALROMQR 417
QY 467 EHATSVKATSAQAWMS 483
Db 418 HRIATVDVKPMVEQAYN 434

RESULT 15
Q938F6 PRELIMINARY; PRT; 603 AA.
AC Q938F6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CYCLODODECANONE MONOOXYGENASE.
GN CDDA.
OS Rhodococcus ruber.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1830;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=SC1;
RX MEDLINE=21475793; PubMed=11591693;
RA Kostichka K., Thomas S.M., Gibson K.J., Nagarajan V., Cheng Q.;
RT "Cloning and Characterization of a Gene Cluster for Cyclododecanone
RL Oxidation in Rhodococcus ruber SCL.";
RL J. Bacteriol. 183:6478-6486(2001).
DR EMBL; AY052630; AAL14233.1; -.
KW Monooxygenase.
SQ SEQUENCE 603 AA; 67536 MW; 29D169ACFE0BD3B2 CRC64;

```

```

Query Match 16.6%; Score 481.5; DB 2; Length 603;
Best Local Similarity 26.4%; Pred. No. 3.3e-31;
Matches 153; Conservative 85; Mismatches 221; Indels 121; Gaps 19;

QY 20 VCAGLSGTSAYVKLRKLRL-NAKIFEGADPGVGVHNNRYPGARVDSPTFPYOLNPIEVM 78
Db 62 IGGFSGLVTAARLRSVESVRIIDKAGDFGVWNNRYPGAMCDTAAWYMPLEEF- 120
QY 79 KDWTSCTRYDQKELLVYHCDKIRGLRKDYV----FGAEYVDARYARDLGTWTVTSA 134
Db 121 -GYMPTKYAHGPEIL---EHCQRI-GKHVDLYDDALFTEVTDLVWQEHQDQWRISTNR 175
QY 135 GHVATAKYLILATGLLHRKHTPALGLADFNKGVIHSSAWHEDF---DAEG-----Q 183
Db 176 GDHFTAQFVGMGTPLHVAQLFGIPGIESFRGSKFHTSRWDYDYGTDGDLGAPMDKLADK 235
QY 184 RVAVIGAGATSIQVQELAKAKADQVTFMFRPSYC-----LPMRQRTMDRNEOTANKAY 238
Db 236 RVAVIGTATVACQVPELAKYCRELYVVOPTSAVDGRGNHPIDENKWFQAIPGQKRW 295
QY 239 PTLFEASRKSRIIGFPVQ-----APSVGIFEVSPQREAYFEE 275
Db 296 LDSFTAIWDGVLTDPSLAIEHEDLVQDGWTALQRMRAAVGSVPTEQYSPENVQRALEE 355

```

```

QY 276 LWERGAFLACQYREVWVDKKANRLVYDFWAKKTRSRIVNPAKRDLMAPLEPPYWG-- 333
Db 356 -----ADDEQMERI-----RAR-VDEIVTDPATAAQLKAWFROM 388
QY 334 TKRSPLESYYEMLDKPSVEIVNLEQSPIVAVTKTGVLLSDGSKRECDTIVLATGDFS- 392
Db 389 CKRPFCHDDYLPAPFNRPNTHLYDTGGKGYERITENGVVVA-GVEYEVDCIVYASGFEFLG 447
QY 393 TCSLTHMGL--KNKHGVDLKEVWKDGISTYMGVFSHGFPNFAFVATAQAFTVLSNGPTII 450
Db 448 TGYTDRAGFDPGTGRDGVKLSHWAQGTRTLHGMHTYGFNLFVLQMQGAALGSNIPHNF 507
QY 451 ETQVDLIADTIAKLEAEHATSVKATSAQAWMSIMI-----A 487
Db 508 VEAARVVAIVDHVJLSTGTSSVETTKAEQAQWVQVLLDHRPLGNPECTPGYNNNEGKPA 567
QY 488 KNEH-----TLFPLTDSWWTGGNIPGKATRALTF 517
Db 568 ELKDRLNVGYPAGSAAFFRMMDHMLAAGSFDG-----LTF 602

```

Search completed: April 27, 2002, 07:57:41
Job time: 697 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:47:14 ; Search time 35.22 Seconds

(without alignments)
599.153 Million cell updates/sec

Title: US-09-882-694A-3

Perfect score: 2902
Sequence: 1 MSATNSRGDCSVACDAIIV.....QICQKVNWDGFDVLHAPC 545

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Swissprot_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|-------|---------------|--------|--------------|--------------------|
| 1 | 802 | 27.6 | 542 | 1 CYMO_ACISP | P12015 acinetobact |
| 2 | 500.5 | 17.2 | 495 | 1 Y892_MYCTU | Q10532 mycobacteri |
| 3 | 427 | 14.7 | 662 | 1 Y41D_RHISN | P55487 rhizobium s |
| 4 | 362.5 | 12.5 | 488 | 1 STCW_EMENI | Q00730 emericeella |
| 5 | 220 | 7.6 | 534 | 1 FM02_MACMU | Q28505 macaca mula |
| 6 | 218 | 7.5 | 534 | 1 FM02_CAVPO | P36366 cavia porce |
| 7 | 216.5 | 7.5 | 532 | 1 FM05_CAVPO | P49109 cavia porce |
| 8 | 216.5 | 7.5 | 532 | 1 FM05_MOUSE | P97872 mus musculu |
| 9 | 213.5 | 7.4 | 532 | 1 FM05_RABIT | Q04799 oryctolagus |
| 10 | 210.5 | 7.3 | 534 | 1 FM02_HUMAN | Q99518 homo sapien |
| 11 | 207.5 | 7.2 | 534 | 1 FM02_RABIT | P17635 oryctolagus |
| 12 | 206.5 | 7.1 | 346 | 1 Y943_MYCTU | P71566 mycobacteri |
| 13 | 202.5 | 7.0 | 532 | 1 FM05_HUMAN | P49326 homo sapien |
| 14 | 200.5 | 6.9 | 530 | 1 FM03_RABIT | P34417 oryctolagus |
| 15 | 198 | 6.8 | 534 | 1 FM03_MOUSE | P97501 mus musculu |
| 16 | 192.5 | 6.6 | 557 | 1 FM04_HUMAN | P31512 homo sapien |
| 17 | 189.5 | 6.5 | 554 | 1 FM04_RABIT | P36367 oryctolagus |
| 18 | 174.5 | 6.0 | 532 | 1 FM01_MOUSE | P50285 mus musculu |
| 19 | 171.5 | 5.9 | 531 | 1 FM03_HUMAN | P31513 homo sapien |
| 20 | 168.5 | 5.8 | 532 | 1 FM01_RAT | P36365 rattus norv |
| 21 | 166.5 | 5.7 | 534 | 1 FM01_RABIT | P17636 oryctolagus |
| 22 | 159 | 5.5 | 531 | 1 FM01_HUMAN | Q01740 homo sapien |
| 23 | 159 | 5.5 | 531 | 1 FM01_PIG | P16549 sus scrofa |
| 24 | 146 | 5.0 | 321 | 1 TRXB_STRCL | P52215 streptomyc |
| 25 | 145.5 | 5.0 | 321 | 1 TRXB_STRCL | Q05741 streptomyc |
| 26 | 137 | 4.7 | 315 | 1 TRXB_MYCPN | P75531 mycoplasma |
| 27 | 135 | 4.7 | 324 | 1 YPDA_BACSU | P50736 bacillus su |
| 28 | 134.5 | 4.6 | 311 | 1 TRXB_HELPY | P58431 helicobacte |
| 29 | 132.5 | 4.6 | 307 | 1 TRXB_TREPA | Q83790 treponema p |
| 30 | 131.5 | 4.5 | 311 | 1 TRXB_HELPJ | Q92118 helicobacte |
| 31 | 128 | 4.4 | 373 | 1 YH6_YEAST | P38866 saccharomyc |
| 32 | 120.5 | 4.2 | 335 | 1 TRXB_MYCTU | P52214 mycobacteri |
| 33 | 117.5 | 4.0 | 458 | 1 TRXB_MYCLE | P46843 mycobacteri |

| | | | | | |
|----|-------|-----|------|--------------|--------------------|
| 34 | 117.5 | 4.0 | 521 | 1 AHFP_ECOLI | P35340 escherichia |
| 35 | 115.5 | 4.0 | 319 | 1 TRXB_LISIN | Q928b5 listeria in |
| 36 | 115.5 | 4.0 | 521 | 1 AHFP_SALTY | P19480 salmonella |
| 37 | 114.5 | 3.9 | 319 | 1 TRXB_LISMO | O32823 listeria mo |
| 38 | 112.5 | 3.9 | 444 | 1 PVDA_BURCE | O51940 burkholderi |
| 39 | 112 | 3.9 | 450 | 1 GSHR_STRTR | O60151 streptococc |
| 40 | 109 | 3.8 | 311 | 1 TRXB_MYCSM | O39973 mycobacteri |
| 41 | 108.5 | 3.7 | 464 | 1 GSHR_SCHPO | P78965 schizosacch |
| 42 | 107 | 3.7 | 315 | 1 TRXB_EUBAC | P50971 eubacterium |
| 43 | 107 | 3.7 | 450 | 1 GSHR_ECOLI | P06715 escherichia |
| 44 | 106 | 3.7 | 4486 | 1 DVH9_HUMAN | Q9nyc9 homo sapien |
| 45 | 104.5 | 3.6 | 409 | 1 TERA_PSESP | P33009 pseudomonas |

ALIGNMENTS

| | | | | | |
|--------------------------|---|-----------|------|---------|--|
| RESULT 1 | | | | | |
| CYMO_ACISP | | | | | |
| ID | CYMO_ACISP | STANDARD; | PRT: | 542 AA. | |
| AC | P12015; | | | | |
| DT | 01-OCT-1989 (Rel. 12, Created) | | | | |
| DT | 01-OCT-1989 (Rel. 12, Last sequence update) | | | | |
| DT | 01-AUG-1991 (Rel. 19, Last annotation update) | | | | |
| DE | Cyclohexanone monooxygenase (EC 1.14.13.22). | | | | |
| OS | Acinetobacter sp. | | | | |
| OC | Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; | | | | |
| OC | Acinetobacter. | | | | |
| OX | NCBI_TaxID=472; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A., AND SEQUENCE OF 1-11. | | | | |
| RC | STRAIN-NCIB 9871; PubMed=3338974; | | | | |
| RX | MEDLINE=88115180; PubMed=3338974; | | | | |
| RA | Chen Y.-C.J., Peoples O.P., Walsh C.T.; | | | | |
| RT | "Acinetobacter cyclohexanone monooxygenase: gene cloning and sequence determination."; | | | | |
| RL | J. Bacteriol. 170:781-789(1988). | | | | |
| CC | -1- CATALYTIC ACTIVITY: Cyclohexanone + NADPH + O(2) -> 6-hexanolide + NADP(4+) + H(2)O. | | | | |
| CC | -1- COFACTOR: FAD. | | | | |
| CC | ----- | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). | | | | |
| CC | ----- | | | | |
| CC | EMBL; M19029; AAA21892.1; . | | | | |
| DR | PIR; A28550; A28550. | | | | |
| DR | InterPro; IPR000960; FMO. | | | | |
| DR | Pfam; PF00743; FMO-like; 1. | | | | |
| KW | Aromatic hydrocarbons catabolism; Oxidoreductase; Monooxygenase; | | | | |
| NADP; Flavoprotein; FAD. | | | | | |
| FT | INIT_MET 0 | | | | |
| FT | NP_BIND 6 18 FAD (BY SIMILARITY). | | | | |
| FT | NP_BIND 176 208 NADP (BY SIMILARITY). | | | | |
| SQ | SEQUENCE 542 AA; 60760 MW; 7ACDD2E5C7BD388F CRC64; | | | | |

Query Match 27.6%; Score 802; DB 1; Length 542;
Best Local Similarity 33.4%; Pred. No. 4.6e-55;
Matches 179; Conservative 96; Mismatches 237; Indels 24; Gaps 8;

| | | | |
|----|----|--|-----|
| QY | 16 | DAIIIGAGLSGTSVYKLR-KLRNKAIFEGADFGGVHWNRYPCARVDSETPFY---- | 70 |
| Db | 7 | DAIIIGGGGGLYAVKRLKDELKVAQADKATDVAGTWNNRYPGALDTETHLICYSW | 66 |
| QY | 71 | ---QLNIPVWKDWTWSCRYPDQKELLSVHHCDKIRGLRKQVYFGAIVVDARYDLGT | 127 |
| Db | 67 | DRELLQSLKIK-----KVVGPDVRKYLQVAKHDLKKSQFNTAVQSAHYNEADAL | 120 |

```
QY 128 WTVKTSAGHVATKAKYLLIATGLLRKHTTPALPLGLADPNKGVIHSSAWHEDFADGQVAV 187
DB 121 WEYTTGYGDKYTARELITLGLLSSAPLNPKINGINQFAGELHHTSRWPDVDSPEGKRVGV 180
QY 188 IGAGATSIQIVQELAKADQVTFMRPSVCLPNRQRTMDRNEQTANKAYPTLFEASRK 247
DB 181 IGTGSGVQVITAVAPLAKHLTVFQSAQVSVPIGNDPLSEEDVKIKDNDYKSLGCMN 240
QY 248 SRIGFPVQAPSGVIFVSPQREAYFEELWER-GAFNFLACQVREVAVMDKKANRLVYDFW 306
DB 241 SALAFALNESTVPAMSVSABERKAVEKAWQTGGFRFMEFETFGDIATNEANIEAQNFI 300
QY 307 AKKTRSRVNPAKRDLMAPEPPYFNGTRKSPLESYDYEMLDKPSVEIVNLEQSPVAVT 366
DB 301 KGKIAEIVKDPALPAQKLPQD---LYAKRPLCDSGYINTFNDRNVRLEDVKANPIVEIT 356
QY 367 KTGVLSDSGSKRCDDTVLATGDFSTGLTHMLKNGKGVLDLKEVWKDGIISYMGVFSH 426
DB 357 ENGCVKLENGDFVELDMICATGDAVDGNTVRMDIOGKNGLAWKDYWKEGPSSYMGVTYN 416
QY 427 GPNPAFFVATAQPTVLSNGPTIETQVDLIADTIKLEAEHATSVEATKSAQEAWSIMI 486
DB 417 NYPNFMVLGPNGP--FTNLPPSIESQVEWISDTIQYTVNNVNSIEATKEABEQWQTC 474
QY 487 AKMNEHTLPLDTSWWTGNGINPGKATRAITFGGIALYEQICOEKVAN--WDGFDV 540
DB 475 ANIAEMTLFPKASQWIFGANIPGKNNTVYFYLGLKEY-RTCASNCKNHAYEGFDI 529

RESULT 2
Y892_MYCTU
ID Y892_MYCTU STANDARD; PRT; 495 AA.
AC Q10532;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE Probable monooxygenase Rv0892 (EC 1.14.13.-).
GN Rv0892 OR MT0916 OR MTCY31.20.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RA complete genome sequence.";
RA Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L. Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouiri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RA laboratory strains.";
RA Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RC -1- COFACTOR: FAD (BY SIMILARITY)
CC -1- SIMILARITY: TO ACINETOBACTER SP. CYCLOHEXANONE MONOOXYGENASE,
CC A.NIDULANS STCW AND TO MAMMALIAN DIMETHYLANILINE MONOOXYGENASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z73101; CA972398.1; -.
DR EMBL: AE006978; AK45162.1; -.
DR TIGR: MT0916; -.
DR TubercuList; RV0892; -.
DR InterPro: IPR000960; FMO.
DR InterPro: IPR000205; NAD_binding.
DR Pfam: PF00743; FMO-like; 1.
KW Hypothetical protein; Oxidoreductase; Monooxygenase; NADP;
KW Flavoprotein; FAD; Complete proteome.
FT NID_BINDING 170 202 NADP (BY SIMILARITY).
SQ SEQUENCE 495 AA; 55039 MW; 08CB55F48674DB44 CRC64;
```

```
Query Match 17.24; Score 500.5; DB 1; Length 495;
Best Local Similarity 28.8%; Pred. No. 1.5e-31;
Matches 143; Conservative 77; Mismatches 217; Indels 59; Gaps 12;

QY 15 CDAT-IVGAGLSGISAVYKRLKRL-NAKIFEGAPDFGVVHNNRYPGARVDSETPFYOL 72
DB 5 CPTVAVVGAGSGMCAITLLSAGITDVCIEKADDVCGTWRDNTYPTGLTCDVPSRLYQ 64

QY 73 NPEVWKDWTWCSCRYPOOKELLSVYHHCOKIRGRKDYVFGAEEVDYARDLGTWTKT 132
DB 65 SFA---KNPNWNTOMFSRGEIQDYLRGIAERYGLRHRIRFGATVVSARF--DDGRWVLR 119

QY 133 SAGHVATKAKYLLIATGLLRKHTTPALPLGLADPNKGVIHSSAWHEDFADGQVAVIGA 192
DB 120 DSGTSTVDFLISATGVLHHPRIPIAGLDDFRGTGVFHSARWHTVPLLRRIAVIGTGS 179

QY 193 TSIQIVQELAKADQVTFMRPSVCLPNRQRTMDRNEQTANKAYPTLFEASRKSR 249
DB 180 TGVOLVCGLAGVAGVNTMFTAQWLPWPNPRYSKLAARVFRHAFPCPLGSLAYKAYSLSF 239

QY 250 IGFVQVQPSVGIPEVSPQREAYFEELWERGAFNFLACQVREVAVMDKKANRLVYDFWAK 309
DB 240 ETFAVALSNFGL-----HRKLVGAVCA 262

QY 310 TRSRVNPAKRDLMAPEPPYFNGTRKSPLESYDYEMLDKPSVEIVNLEQSPVAVTGTG 369
DB 263 SLRRVRDPLRLRALTPDYEPM---CKRLVMSGGFYRAIQDDVELVT---AGIDHVEHRG 316

QY 370 VLLSDGSKRCDDTVLATGDFSTGLTHMLKNGKGVLDLKEVWKDGIISYMGVFSHGFP 429
DB 317 IVTDDGVLHEVDVIVLATGDFDSH-AFPRPMLTGRDGIIRDVWQDGPFAHQTVAIQGP 375

QY 430 NAFVATAQAPTVLSNGP--TILETQVDLIADTIKLEAEHATSVEATKSAQEAWSIMI 487
DB 376 NFMMLGPHSP---VGNFPLTAVAESQAEHIVQWIKRHRHGEFDTMEPKSAATEAYNTVLR 433

QY 488 KMNEHTLPLDTSWWTGNGINPGKATRAITFGGIALYEQICOEKVAN--WDGFDV 502
DB 434 AAMPNTVWTGCDSWY 449
```

```
RESULT 3
Y41D_RHISN
ID Y41D_RHISN STANDARD; PRT; 662 AA.
AC P55487;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Probable monooxygenase Y41D (EC 1.14.13.-).
GN Y41D.
OS Rhizobium sp. (strain NGR234).
OC Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
```

| RESULT | 4 | STC | EMENI | STANDARD | PRT | 488 AA |
|--------|--|-----------------|-------|----------|-----|--------|
| OC | Rhizobiaceae; Rhizobium. | NCBI_TaxID=394; | | | | |
| OC | NCBI_TaxID=394; | | | | | |
| OC | SEQUENCE FROM N.A. | | | | | |
| OC | MEDLINE=97305956; PubMed=9163424; | | | | | |
| OC | Freiburg C.A., Fellay R., Broughton W.J., Rosenthal A., | | | | | |
| OC | Perret X., | | | | | |
| OC | "Molecular basis of symbiosis between Rhizobium and legumes." | | | | | |
| OC | Nature 387:394-401(1997). | | | | | |
| OC | -!- COFACTOR: FAD (BY SIMILARITY). | | | | | |
| OC | -!- SIMILARITY: TO ACINETOBACTER SP. CYCLOHEXANONE MONOOXYGENASE, | | | | | |
| OC | M. TUBERCULOSIS RV0892, A. NIDULANS STCW AND TO MAMMALIAN | | | | | |
| OC | DIMETHYLANILINE MONOOXYGENASES. | | | | | |
| OC | ----- | | | | | |
| OC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | | | |
| OC | between the Swiss Institute of Bioinformatics and the EMBL Outstation | | | | | |
| OC | the European Bioinformatics Institute. There are no restrictions on its | | | | | |
| OC | use by non-profit institutions as long as its content is in no way | | | | | |
| OC | modified and this statement is not removed. Usage by and for commercial | | | | | |
| OC | entities requires a license agreement (See http://www.isb-sib.ch/announcement/ | | | | | |
| OC | or send an email to license@isb-sib.ch). | | | | | |
| OC | ----- | | | | | |
| OC | EMBL: AE000078; AAB91699.1; - | | | | | |
| OC | InterPro: IPR000960; FMO. | | | | | |
| OC | InterPro: IPR000205; NAD-binding. | | | | | |
| OC | Pfam: PF00743; FMO-like; 1. | | | | | |
| OC | Hypothetical protein; Oxidoreductase; Monooxygenase; NADP; | | | | | |
| OC | Flavoprotein; FAD; Plasmid. | | | | | |
| OC | NP_BIND 303 335 NADP (BY SIMILARITY). | | | | | |
| OC | SEQUENCE 662 AA; 75109 MW; 524689EA62AA76CB CRC64; | | | | | |
| OC | ----- | | | | | |
| OC | Query Match 14.7%; Score 427; DB 1; Length 662; | | | | | |
| OC | Best Local Similarity 25.9%; Pred. No. 1.2e-25; | | | | | |
| OC | Matches 131; Conservative 87; Mismatches 208; Indels 80; Gaps 15; | | | | | |
| OC | ----- | | | | | |
| OC | 18 IIIVAGLSGISAVYKLRKLRNAKIFEGAPDGGVWVWNNRYPGARVDSETPFYQLNIP 77 | | | | | |
| OC | IIIVAGLSGISAVYKLRKLRNAKIFEGAPDGGVWVWNNRYPGARVDSETPFYQLNIP 77 | | | | | |
| OC | 138 LIIGAGMSVAAAIIRLGGISIIQVKEQDSTGGVWVWNNRYPGARVDSETPFYQLNIP 193 | | | | | |
| OC | LIIGAGMSVAAAIIRLGGISIIQVKEQDSTGGVWVWNNRYPGARVDSETPFYQLNIP 193 | | | | | |
| OC | 78 WKDWTWSCRYPDQKELLSYVHHCKIRGLRDVYFGAIVDARYARDLGTW 133 | | | | | |
| OC | WKDWTWSCRYPDQKELLSYVHHCKIRGLRDVYFGAIVDARYARDLGTW 133 | | | | | |
| OC | 194 FASGNWSTFPDQKELLSYVHHCKIRGLRDVYFGAIVDARYARDLGTW 253 | | | | | |
| OC | FASGNWSTFPDQKELLSYVHHCKIRGLRDVYFGAIVDARYARDLGTW 253 | | | | | |
| OC | 134 AGHVATAKYLLIATGLLRKHTTPALPGADNGKVIHSSAWHEDFDAGQVAVIGAGAT 193 | | | | | |
| OC | AGHVATAKYLLIATGLLRKHTTPALPGADNGKVIHSSAWHEDFDAGQVAVIGAGAT 193 | | | | | |
| OC | 254 TEETLVNIVLAVSGVGTTPKPNLNSGLRNFDPGVVHTSKWDPVALDGRVAVIGNGAS 313 | | | | | |
| OC | TEETLVNIVLAVSGVGTTPKPNLNSGLRNFDPGVVHTSKWDPVALDGRVAVIGNGAS 313 | | | | | |
| OC | 194 SIQIVQELAKKADQVMTMRPSYCLPQRQRTMRNEQTAKVYPTLFEASRSRIGFP 253 | | | | | |
| OC | SIQIVQELAKKADQVMTMRPSYCLPQRQRTMRNEQTAKVYPTLFEASRSRIGFP 253 | | | | | |
| OC | 314 AMQVPAIADRYGALTIFQR-----SRQ-----WVAPP-----KFOKP 347 | | | | | |
| OC | AMQVPAIADRYGALTIFQR-----SRQ-----WVAPP-----KFOKP 347 | | | | | |
| OC | 254 VQAPSVGIFEVSPQEQREAYFELWGRGAFNPLACQYREVWYDKKANRLVDFWAKKTR 312 | | | | | |
| OC | VQAPSVGIFEVSPQEQREAYFELWGRGAFNPLACQYREVWYDKKANRLVDFWAKKTR 312 | | | | | |
| OC | 348 VPEWQFLREVP-----HYEWLYRLSLWYIDSEVHEAKQKPA-----WHPDKSV 395 | | | | | |
| OC | VPEWQFLREVP-----HYEWLYRLSLWYIDSEVHEAKQKPA-----WHPDKSV 395 | | | | | |
| OC | 313 RIYNPAKR-----DLMAPLEPPYFMTGRKSPLESYEMLDKPSVEIVN- 356 | | | | | |
| OC | RIYNPAKR-----DLMAPLEPPYFMTGRKSPLESYEMLDKPSVEIVN- 356 | | | | | |
| OC | 396 NAVNDRDREAYTQYIEGQLAGRPDLIAKIVPSYPPGKRMELDNGVYRTLLKPHVTLVDG 455 | | | | | |
| OC | NAVNDRDREAYTQYIEGQLAGRPDLIAKIVPSYPPGKRMELDNGVYRTLLKPHVTLVDG 455 | | | | | |

```
Db 261 -GAF-----YRDSMAQKASAFF-----RORMASIIKDDRLAKGFTPTFGGCRRI 305
Qy 338 PLESYYEMLQPSVEIVNLQSPITAVTKTGVLSDGSKRECDTIVLATGFD-SPTGSL 396
Db 306 TPGDPYMHAIQANVDV---HFTVASCTEDGIVGADGIERLVITVCASGDFUNTYRPF 362
Qy 397 THMGLKNGHVDLKEVKNQDGYSTYMGVSHGFPNFAFFVATAQAAPTIVLSNGPTIETQVDL 456
Db 363 PIIG---RRGVDLRDWKNTPEAYLGLAVPMPN-VITFTGSPWPIQNG-----407
Qy 457 IADTIKLEAHATSVKATSKAQ-----EAWG---IMIAKMHNEH 492
Db 408 --SVNAPLHVSEYATQFLKMKQENIRAWAPROQITDRFNEH 448

RESULT 5
FM02_MACMU STANDARD; PRT; 534 AA.
AC Q28505;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Dimethylalanine monooxygenase [N-oxide forming] 2 (EC 1.14.13.8)
DE (Pulmonary flavin-containing monooxygenase 2) (FMO 2) (Dimethylalanine
  oxidase 2) (FMO 1B1).
DN FMO2.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheciae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97214641; PubMed=9061021;
RX Yueh M.F., Krueger S.K., Williams D.E.;
RT "Pulmonary flavin-containing monooxygenase (FMO) in rhesus macaque:
  expression of FMO2 protein, mRNA and analysis of the cDNA.";
RL Biochem. Biophys. Acta 1350:267-271(1997).
CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN THE OXIDATIVE METABOLISM OF
  A VARIETY OF XENOBIOTICS SUCH AS DRUGS AND PESTICIDES.
CC -1- CATALYTIC ACTIVITY: N,N-dimethylalanine + NADPH + O(2) -> N,N-
  dimethylalanine N-oxide + NADP(+) + H(2)O.
CC -1- COFACTOR: FAD AND MAGNESIUM ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Microsomal.
CC -1- SIMILARITY: BELONGS TO THE FMO FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
DR EMBL; U59453; AAB02939.1; .
DR InterPro; IPR000960; FMO.
DR Pfam; PF00743; FMO-like; 1.
DR PRINTS; PR00370; FMOXYGENASE.
KW Oxidoreductase; Monooxygenase; NADP; Flavoprotein; FAD; Microsome;
KW Transmembrane; Multigene family; Acetylation; Magnesium.
FT INIT_MET 0 0 BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT NP_BIND 8 13 FAD (ADP PART) (POTENTIAL).
FT NP_BIND 190 195 NADP (POTENTIAL).
SQ SEQUENCE 534 AA; 60830 MW; B77D6F03B07CDB59 CRC64;

Query Match 7.68; Score 220; DB 1; Length 534;
Best Local Similarity 21.88; Pred. No. 1.3e-09;
Matches 103; Conservative 71; Mismatches 194; Indels 104; Gaps 15;
Qy 19 IVGAGLSGISAVYKLRKLRLNAKIFEGAPDFGGVWHNRYPGARVDSETPFYQLNIFEVW 78
```

```
Db 6 VIGAGVSLISLKKCVDEGLEPTCFERTEDTGGVW---RFEKVEDGRASIYQVNTS 62
Qy 79 KDWIWSCRYDOKELLSYVHHCKIRGLR-----KDYVFGAEVVDARYARDL--G 126
Db 63 KEMSCFSDFPMPEDFPNHLNSKLEFYRFPAAKFDLLKYIQFOTTVLSVRKCDPSSG 122
Qy 127 TWTVKT-SAGHVATKY--LILATGLLHRKHTP--ALPLGLADPNKGVIHSSAMHEDFAE 181
Db 123 QKWVVTQSNKEQSADVDVAVVCTGHHFLPHILPLKSPGIERFKQGFHSRQYKHPDGE 182
Qy 182 GORVAVIGAGATSIQIVQELAKYADVTMWRPESYCLPMRQRTMRNEQTAKKAYPTL 241
Db 183 GKRLIVGMNGSGSDIAVELSKSAQVFISTRHGTWV-----220
Qy 242 FEASRSRIGFPVQAPSVGIFEVSPEQREAYFEELWERGAFNFLACQYREVMVDKKANRL 301
Db 221 ---SRVSEDCYP-----WDSVFHTFRSMLRNVLPTVTKWMIQOMNQ- 261
Qy 302 VYDFWAKKTRSRIVNPAKRLMAPLEPPYWGFKRSPLSDYEMLDKPSVEIVNLQSP 361
Db 262 ----WFNHN-----YGLEPQNKYIMKEPVNDNDVPSRLCGGALKV---RST 301
Qy 362 IVAVTKTGVLLSDGSKRE-CDTIVLATGFSFTGSLTHMGLKNGHGYDLKREVMKDGISTY 420
Db 302 VKELTETSATFEDGTVEENIDVIFATGY-----SFSFPFLED---SLVKVENNMVSLY 352
Qy 421 MGVFSHGFPNFAFFVATAQAAPTIVLSNGPTIETQVDLLIADTIKLEAHATSV 472
Db 353 KYIFPAHLERKSTACTIGLIPLASIFPT-----AELQARWVTRV 391

RESULT 6
FM02_CAVPO STANDARD; PRT; 534 AA.
AC P36366; Q05194;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Dimethylalanine monooxygenase [N-oxide forming] 2 (EC 1.14.13.8)
DE (Pulmonary flavin-containing monooxygenase 2) (FMO 2) (Dimethylalanine
  oxidase 2) (FMO 1B1).
DN FMO2 OR FMO-2.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=HARTLEY; TISSUE=Lung;
RX MEDLINE=93306345; PubMed=1306120;
RX Nkbbakht K.N., Lawton M.P., Philpot R.M.;
RT "Guinea pig or rabbit lung flavin-containing monooxygenases with
  distinct mobilities in SDS-PAGE are allelic variants that differ at
  only two positions.";
RT Pharmacogenetics 2:207-216(1992).
RL -1- FUNCTION: THIS PROTEIN IS INVOLVED IN THE OXIDATIVE METABOLISM OF
  A VARIETY OF XENOBIOTICS SUCH AS DRUGS AND PESTICIDES.
CC -1- CATALYTIC ACTIVITY: N,N-dimethylalanine + NADPH + O(2) -> N,N-
  dimethylalanine N-oxide + NADP(+) + H(2)O.
CC -1- COFACTOR: FAD AND MAGNESIUM ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Microsomal.
CC -1- TISSUE SPECIFICITY: LUNG.
CC -1- POLYMORPHISM: THERE ARE TWO ALLELIC FORMS (A AND B).
CC -1- SIMILARITY: BELONGS TO THE FMO FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
```


| | |
|---|--|
| CC | EMBL; L10037; AAB59631.1; - |
| DR | InterPro; IPR000960; FMO. |
| DR | Pfam; PF00743; FMO-like; 1. |
| DR | PRINTS; PR00370; FMOXYGENASE. |
| KW | Oxidoreductase; Monooxygenase; NADP; Flavoprotein; FAD; Microsome; |
| FT | Transmembrane; Multigene family; Acetylation; Polymorphism; Magnesium. |
| FT | INIT_MET 0 0 BY SIMILARITY. |
| FT | MOD_RES 1 1 ACETYLATION (BY SIMILARITY). |
| FT | NP_BIND 8 13 FAD (ADP PART) (POTENTIAL). |
| FT | NP_BIND 190 195 NADP (POTENTIAL). |
| FT | VARIANT 521 521 V -> L (IN B FORM). |
| FT | VARIANT 529 529 S -> F (IN B FORM). |
| FT | SEQUENCE 534 AA; 60810 MW; 22C6A0A06EB59955 CRC64; |
| | |
| | Query Match 7.5%; Score 218; DB 1; Length 534; |
| | Best Local Similarity 21.6%; Pred. No. 1.9e-09; |
| | Matches 104; Conservative 74; Mismatches 181; Indels 122; Gaps 17; |
| Qy | 19 IVGAGLSGISAVYKRLRLNAKIFEGAPDGGYWHNNRYPGARVDSPTFPYQLNIPW 78 |
| Db | 6 VIGAVSSLSLKCCEBEGLEPTCFERTEDIGGLW---RPFENVEDGRASIYKSVITNTS 62 |
| | |
| Qy | 79 KDWTSRCRYPDQKELLSVHHCDKIRGLR-----KDYFGAEVVDARYARDL---G 126 |
| Db | 63 KEMSCFSDPMPEDFPNPLHNSKLLLEYFLRFAKFDLLKYIQFQTTLTVTKKHPDFSSG 122 |
| | |
| Qy | 127 TWTVKT-SAGHVATKY--LILATGLLHRKHTP--ALPLADFNKGKVTSSAWHEDFAE 181 |
| Db | 123 QMEVVTQSGKEQSAVDANVWCVSGHHLLPHIPLKSFPEIERFGYFHSQYKHPAGFE 182 |
| | |
| Qy | 182 GORVAVIGAGATSIIQVQELAKKADQVTFMRMRPSCYCLPMRQRTMDRNEQTAWKAYPTL 241 |
| Db | 183 GKRLIVIGIGNSADISSELSKNAQVFISTRNGSWN----- 220 |
| | |
| Qy | 242 FEARKSRIGFPVQAPSGVGEFSEPEQREAYFEELWERGAFNLCQYREVWVKKANRL 301 |
| Db | 221 ---SRISDGYP-----WDMVFHTRFKSLRNILPRTVSKWMEQOLAR- 261 |
| | |
| Qy | 302 VYDFWAKTRSRIVNPAKROLMAPLEPPYWFCTKRSPLSDSYEMLDKPSVEIVNLEQSP 361 |
| Db | 262 ----W-----FNHAYS-----LEPNKYLMKEPIINDLPSRLIYGAVKV---KSR 301 |
| | |
| Qy | 362 IVAVYTKTVLLSDGS-KRECDTIVLATGDFSPTGSLTHMGLNKGVDLKEVWKDGISTY 420 |
| Db | 302 VTQLTETSALEDGTVEEDIDVVFATGCTSFPPLESLSYKIEHM----- 348 |
| | |
| Qy | 421 MGVFSGHPNFAFFVATAQAPTVLSNG-----PTTIETQVDLIADTIKLEAEHATS 471 |
| Db | 349 VSLKYMEFP-----PQLEKPTLTCMGLIQLPLGSIFPTV-----ELQARWATR 390 |
| | |
| Qy | 472 V 472 |
| Db | 391 V 391 |
| | |
| RESULT 7 | |
| FMO5_CAVPO | STANDARD; PRT; 532 AA. |
| ID -FMO5_CAVPO | AC P49109; |
| DT 01-FEB-1996 (Rel. 33, Created) | |
| DT 01-FEB-1996 (Rel. 33, Last sequence update) | |
| DT 01-FEB-1996 (Rel. 33, Last annotation update) | |
| DE Dimethylalanine monooxygenase [N-oxide forming] 5 (EC 1.14.13.8) | |
| DE (Hepatic flavin-containing monooxygenase 5) (FMO 5) (Dimethylalanine oxidase 5). | |
| DE FMO5. | |
| OS Cavia porcellus (Guinea pig). | |
| OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia. | |
| OX NCBI_TaxID=10141; | |
| RN [1] | |

[illegible]

QY 473 -----BATKSAQEAWSIMIAK-----MNEHTL 494
DB 395 LKTLPSQSEMMBITRAQEE-----IAKRYVDSQRHTI 427

RESULT 8

FM05_MOUSE
ID FM05_MOUSE STANDARD; PRT: 532 AA.
AC P97872;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Dimethylalanine monooxygenase [N-oxide forming] 5 (EC 1.14.13.8)
DE (Hepatic flavin-containing monooxygenase 5) (FMO 5) (Dimethylalanine
oxidase 5).
GN FM05.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1; TISSUE=Liver;
RX MEDLINE=98241973; PubMed=9580872;
RA Cherrington N.J., Falls J.G., Rose R.L., Clements K.M., Philpot R.M.,
RA Levi P.E., Hodgson E.;
RT "Molecular cloning, sequence, and expression of mouse flavin-
containing monooxygenases 1 and 5 (FM01 and FM05).";
RL J. Biochem. Mol. Toxicol. 12:205-212(1998).
CC -!- FUNCTION: IN CONTRAST WITH OTHER FORMS OF FMO IT DOES NOT SEEM TO
BE A DRUG-METABOLIZING ENZYME (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: N,N-dimethylalanine + NADPH + O(2) = N,N-
dimethylalanine N-oxide + NADP(+) + H(2)O.
CC -!- COFACTOR: FAD.
CC -!- SUBCELLULAR LOCATION: Mitochondrion.
CC -!- SIMILARITY: BELONGS TO THE FMO FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DB EMBL; U90535; AAB50013.1; -;
DR MGD; MGI:1310004; Fmo5.
DR InterPro; IPR000960; FMO.
DR Pfam; PF00743; FMO-like; 1.
DR PRINTS; PR00370; FMOXYGENASE.
KW Oxidoreductase; Monooxygenase; NADP; Flavoprotein; FAD; Mitochondrion;
KW Transmembrane; Multigene family; Acetylation.
FT INIT_MET 0 BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT NP_BIND 9 14 FAD (ADP PART) (POTENTIAL).
FT NP_BIND 191 196 NADP (BY SIMILARITY).
SQ SEQUENCE 532 AA; 59898 MW; B1749652B30F04D5 CRC64;

Query Match 7.5%; Score 216.5; DB 1; Length 532;
Best Local Similarity 21.68; Pred No. 2.5e-09;
Matches 108; Conservative 77; Mismatches 192; Indels 123; Gaps 20;

QY 19 IVGAGLSGISAVYKLRKLRNNAKIFGAPDFGVVHNRNYFCARVDSETPFYQLNPIVW 78
DB 7 VIGAGSLTCKICLERGLEPVCFVSCDGLGLW---RFQEAPEGRASIVQSVVNTS 63
QY 79 KDWTSRCYRPODKELLSVHHCDKIRGLR-----KDYFGEVVDARYADRL---G 126
DB 64 KEMCFSDYIPDHPYVNMHNSQVLEVFMYAKFEDLLKYTFQFTVCSVKKQPDFSTSG 123
QY 127 TWTVTSA---GHVATARYLILATGLLRKHTP--ALPLGADFNCKVIHSSAHWDFDAE 181

DB 124 QMQVVTECEGKQOVDFGVGLVCTGHHTDAHLPLESFCPIERFKYKYPHSRDYKNPVET 183
QY 182 QORVAVIGAGATSIIQVQELAKKADQVTFMFRPSPCYCLPMRQRTWDRNEQTANKAYPTL 241
DB 184 GKRIVIVIGNSGGDLAVEISHTAKQVFLSTRRGAWIL----- 221
QY 242 FEASRKRIGFPVO--APSVGIFEVS---PEOREAYEELWERGAFNFLACQYREVMD 295
DB 222 ---NRVKGHGYPIDLLLSRIMYLSRIGCGSLKNKYWE----- 257
QY 296 KKAN-RLVYDFWAKTRSRIVNPAKRDLMAPLEPPYWFCTKRSPLESYEMLDKPSVET 354
DB 258 KQNRQFDHEMFLKPKHRALS-----QHPVND-----DLPNRIL 293
QY 355 VNL--EQSPIVAVTKTGVLVLLSDGSKRE--CDTIVLATGDSFTGSLTHMGLKNKHGVDLKE 411
DB 294 AGLVKVGNVKEFTETAIFEDSGREDGIDVIVFATGY-SFAFFLE-----DSVK 343
QY 412 VKWDGISTYMGVFSNGFPNFAFVATAOAPTVLNSGPTIIETQVDLIADTIAKLEAFHATS 471
DB 344 VVKNVSLYKKVFP---PN-----LEKPTLAIG--LIQPLGAIM--PISELQGRWATQ 390
QY 472 VEATKSAQEAWSIMIAKME 491
DB 391 VFGLKGLKLPQSQSEMAEINK 410

RESULT 9

FM05_RABIT
ID FM05_RABIT STANDARD; PRT: 532 AA.
AC Q04799;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Dimethylalanine monooxygenase [N-oxide forming] 5 (EC 1.14.13.8)
DE (Hepatic flavin-containing monooxygenase 5) (FMO 5) (Dimethylalanine
oxidase 5) (FMO 1C1) (FMO form 3).
GN FM05.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93252844; PubMed=8486656;
RA Atta-Asafo-Adjei E., Lawton M.P., Philpot R.M.;
RT "Cloning, sequencing, distribution, and expression in Escherichia
coli of flavin-containing monooxygenase 1C1. Evidence for a third
gene subfamily in rabbits.";
RT J. Biol. Chem. 268:9681-9689(1993).
RN [2]
RP SEQUENCE OF 2-513.
RC STRAIN=NEW ZEALAND WHITE; TISSUE=Liver;
RX MEDLINE=94190897; PubMed=8142375;
RA Ozols J.;
RT "Isolation and structure of a third form of liver microsomal flavin
monooxygenase.";
RL Biochemistry 33:3751-3757(1994).
CC -!- FUNCTION: IN CONTRAST WITH OTHER FORMS OF FMO IT DOES NOT SEEM TO
BE A DRUG-METABOLIZING ENZYME.
CC -!- CATALYTIC ACTIVITY: N,N-dimethylalanine + NADPH + O(2) = N,N-
dimethylalanine N-oxide + NADP(+) + H(2)O.
CC -!- COFACTOR: FAD.
CC -!- SUBCELLULAR LOCATION: Mitochondrion.
CC -!- TISSUE SPECIFICITY: KIDNEY AND LIVER.
CC -!- SIMILARITY: BELONGS TO THE FMO FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way


```
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 37.2 kDa protein RV0943c.
GN RV0943C OR MT0969 OR MTCY10D7.31.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K.K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayan L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bisht W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licensed@isb-sib.ch).
CC -----
DR EMBL: Z79700; CAB01978.1; -.
DR EMBL: AF006982; AAK45217.1; ALT_INIT.
DR TIGR: WT0969; -.
DR TubercuList; RV0943c; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 346 AA; 37205 MW; F51A5815B5A36136 CRC64;
-----
Query Match              7.1%; Score 206.5; DB 1; Length 346;
Best Local Similarity    21.1%; Pred. No. 8,3e+09;
Matches      71; Conserved   39; Mismatches   95; Indels 131; Gaps       9;

QY 171 SSAWHEDFAEGORVAVIGAGATSIQIQVLAKKADQVTMPMRPSYCLPQRQTMDRNE 230
Db |::| ||||| ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
31 SAAWDRDFDPAGRIAVGVGTDAAAHHYSIRLSAASAVTVTGQP----- 75

QY 231 QTAWKAYIPTLFASKRSKRIGFFVPQSPVGIFEVSPEQREAYFEELWERGFNFLACQYR 290
Db |:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
76 -----RRVTVGP-----LMTTRAKRWL----- 93

QY 291 EVMWDKKANRLVDYNWAKTKTSRIIVNPDAKDLMAPLEPPYWGFKRSPLESDDYEMLDKP 350
Db |:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
94 -----RRRTGAEPHA----- 103

QY 351 SVEIVNLQSPIVAVNTKTGLLLSDGSKRKCDTI VLATGFDSTFSGSLTHMGLKNKHGYDLK 410
Db |:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
104 ----VAWATAIIDLITSSGIIRTSDGEHPHVDAIIYGTF-AIAQQVGSDQTLVGAGGVYIR 158

QY 411 EVWKDGISTYNGVFSHGFENAFVAAQAAPT VLNSNGPTIITETOVDDLADIATIAKLEABHAT 470
```

```

      : | ||: |:|| ||||| ||: || | : : : |:|
Db 159 QAWDDGMEPYLGVAHGFPPNYFFI-----TQPD-TAAQAQCVCVKMLMRTASR 207
QY 471 SVEATKSAQGAWSIMIAKMNEHT-LPPLTDSDWWTGG 505
Db 208 RIEVRSQQVF-----NERAQLKPAQPHROTGG 236
      : | ||: |:|| ||||| ||: || | : : : |:|

RESULT 13
FM05_HUMAN STANDARD; PRT; 532 AA.
AC P49326;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Dimethylaniline monooxygenase [N-oxide forming] 5 (EC 1.14.13.8)
DE (hepatic flavin-containing monooxygenase 5) (PMO 5) (Dimethylaniline
DE oxidase 5).
DS FMO5.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=95177663; PubMed=7872795;
RA Overby L.H., Buckpitt A.R., Lawton M.P., Atta-Asafo-Adjei E.,
RA Schulze J., Philpot R.W.;
RT "Characterization of flavin-containing monooxygenase 5 (FMO5) cloned
RT from human and guinea pig: evidence that the unique catalytic
RT properties of FMO5 are not confined to the rabbit ortholog.";
RL Arch. Biochem. Biophys. 317:275-284(1995).
RN [2].
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta, and Liver;
RA Dolphin C.T., Povey S., Shephard E.A., Smith R.L., Phillips I.R.;
RL SUBMITTED (JAN-1995) to the EMBL/GenBank/DDBJ databases.
CC -1- FUNCTION: IN CONTRAST WITH OTHER FORMS OF FMO IT DOES NOT SEEM TO
CC BE A DRUG-METABOLIZING ENZYME.
CC -1- CATALYTIC ACTIVITY: N,N-dimethylaniline + NADPH + O(2) = N,N-
CC dimethylaniline N-oxide + NADP(+) + H(2)O.
CC -1- COFACTOR: FAD.
CC -1- SUBCELLULAR LOCATION: Microsomal.
CC -1- TISSUE SPECIFICITY: FETAL AND ADULT LIVER.
CC -1- SIMILARITY: BELONGS TO THE FMO FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collabor-
CC ation between the Swiss Institute of Bioinformatics and the EMBL Outstat-
CC ing Centre at the European Bioinformatics Institute. There are no restrictions on
CC the use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commer-
CC cial entities requires a license agreement (See http://www.isb-sib.ch/anno-
CC te.html or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L37080; AAA67849.1; -
CC EMBL: Z47553; CAA87633.1; -
CC MIM: 603957; -
CC InterPro: IPR000960; FMO.
CC Pfam: PF00743; FMO-like; 1.
CC PRINTS: PR00370; FMOXYGENASE.
CC Oxidoreductase; Monooxygenase; NADP: Flavoprotein; FAD; Microsome;
KW Transmembrane; Multigene family; Acetylprote.
FT INIT_MET 0 0 BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT NP_BIND 9 14 FAD (ADP PART) (POTENTIAL).
FT NP_BIND 191 196 NADP (BY SIMILARITY).
FT CONFLICT 350 350 S -> P (IN REF. 2).
SQ SEQUENCE 532 AA; 60089 MW; D4D9F092E52A56D1 CRC64;

```



```
Db 250 -----ISDWVVKOMNAKFKHENYSLM-PLN-----GTLRKPEVND-----DLP 288
QY 351 S-----VEIVNLEOSPPIVAVTKTVLLSDGSKREC-DTIVLATGSDFTSGTSLTHMGLKNKHG 406
Db 289 ARILCTGVSIKPN-VKEFTETSAIFEDGTVEAIDSVIFATGYGYAYFPFLDSDSIKSEN- 346
QY 407 VDLKEWKDGISYMGVFSHGFPNAFFVATAQAAPTIVLSNGPTTIETQVDLIADTIKLEA 466
Db 347 -----NKVTLFKGIF-----PPQLEKPTMAVIGLVOSLGA 376
QY 467 EHATSVEATKSAQFAWSIMIAKNEHTLFPLTD 499
Db 377 AIPPT-----DLQARWAAQVIK-GTCTLPVKD 403

RESULT 15
FMO3.MOUSE
ID FMO3.MOUSE STANDARD; PRT; 534 AA.
AC P97501;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Dimethylalanine monooxygenase [N-oxide forming] 3 (EC 1.14.13.8)
DE (Hepatic flavin-containing monooxygenase 3) (FMO 3) (Dimethylalanine
DE oxidase 3).
DE FMO3.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CD-1; TISSUE=Liver;
RX MEDLINE=98008021; PubMed=9344459;
RA Falls J.G., Cherrington N.J., Clements K.M., Philpot R.M., Levi P.E.,
RA Rose R.L., Hodgson E.;
RT "Molecular cloning, sequencing, and expression in Escherichia coli of
RT mouse flavin-containing monooxygenase 3 (FMO3); comparison with the
RT human isoform.";
RL Arch. Biochem. Biophys. 347:9-18(1997).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE OXIDATIVE METABOLISM OF
CC A VARIETY OF XENOBIOTICS SUCH AS DRUGS AND PESTICIDES. FORM II
CC N-OXYGENATES PRIMARY ALIPHATIC ALKYLAMINES AS WELL AS SECONDARY
CC AND TERTIARY AMINES (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: N,N-dimethylalanine + NADPH + O(2) -> N,N-
CC dimethylalanine N-oxide + NADP(+) + H(2)O.
CC -!- COFACTOR: FAD (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Microsomal.
CC -!- SIMILARITY: BELONGS TO THE FMO FAMILY.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: U87147; AAB47541.1; -
CC MGD: MGI:1100496; Fmo3.
CC InterPro: IPR000960; FMO.
CC Pfam: PF00743; FMO-like; 1.
CC PRINTS: PR00370; FMOXYGENASE.
CC Oxidoreductase; Monooxygenase; NADP; Flavoprotein; FAD; Microsome;
KW Transmembrane; Multigene family.
FT NP_BIND 9 14 FAD (ADP PART) (POTENTIAL).
FT NP_BIND 191 196 NADP (POTENTIAL)..
SQ SEQUENCE 534 AA; 60516 MW; F72F7993C01AF9C9 CRC64;
```

Query Match 6.8%; Score 198; DB 1; Length 534;
Best Local Similarity 21.2%; Pred. No. 6.9e-08;

```
Matches 103; Conservative 69; Mismatches 183; Indels 130; Gaps 19;
QY 19 IVGAGLSGISAVYKLRKLRNAKIFEGAPDFGGVWHNRYPGARVDSETPFQOLNIPW 78
Db 7 IIGAGVSGLAAIRSCLEGLEPTCFERSDDVGGGLW---KFSDHIEGRASISQSVFTNSS 63
QY 79 KDMTWSCRYPDQKELLISVHHCD-----KIRGLRKDVYFGAIVVDARYARDL---G 126
Db 64 KEMACFPDPFPYDDFPNPMHHSKLQEIYTSFAKERNLLKYIQFETPVTISNKCPSNFTTG 123
QY 127 TWTVKTSA-GHVATAKY--LILATGLLRHKHP--ALPGLADFNKVIHSSAWHEDFDAE 181
Db 124 KWEVTTKHKGKETAVFDATMICSGHHIFPHVPKDSFPGLNRFKGCFCFHSRDYKEFGIWK 183
QY 182 GORVAVIGAGATSIQIVQELAKKADQVTMFMRRPSYCLPMRQRTMDRNEQTAKAYIP-- 239
Db 184 GKRVLVIGLNSGCCDIAAELSHVAOKVTISSRSGSWM-----SRVWDDGYPWD 232
QY 240 ---TLFEASRSKRIGFPVQAPSVGIFEVSPQREAYFEELWERGAFNACQYREVMVD 295
Db 233 MVVLTRFQTLKNNL-----PTA----- 250
QY 296 KKANRLVYDFW-AKTRSRIVN-----PAKRDLMAPLEPPYWFGTGKRSPLESDDYEMLD 348
Db 251 -----ISDWWVTROMNARFKHENYGLVPLNRTL-----RKEPVFND--ELPA 290
QY 349 KPSVEIVNLEOSPPIVAVTKTVLLSDGSKREC-DTIVLATGSDFTSGTSLTHMGLKNKHG 407
Db 291 RILCGMVTIKPN-VKEFTETSAIFEDGTVEAIDCVIFATGYGYAYFPFLDSDSIKSRN-- 347
QY 408 DLKEWKDGISYMGVFSHGFPNAFFVATAQAAPTIVLSNGPTTIETQVDLIADTIKLEA 467
Db 348 -----NEVTLYKGVFP---PQLEKPTMAVIGLVQSLGATI-----PITDQAR 387
QY 468 HATSV 472
Db 388 WAAQV 392
```

Search completed: April 27, 2002, 07:59:37
Job time: 743 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 05:56:13 ; Search time 68.15 Seconds
(without alignments)
768.433 Million cell updates/sec

Title: US-09-882-694A-3
Perfect score: 2902
Sequence: 1 MSATNSRSGDCSVACDAIIV.....QICQEKVANWDGFDVLHAPC 545

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------|---------------------|
| 1 | 895.5 | 30.9 | 549 | JC7158 | steroid monooxygen |
| 2 | 838 | 28.9 | 556 | C87472 | steroid monooxygen |
| 3 | 802 | 27.6 | 542 | A28550 | cyclohexanone mono |
| 4 | 786 | 27.1 | 543 | E87474 | steroid monooxygen |
| 5 | 550.5 | 19.0 | 491 | F83383 | probable flavin-bi |
| 6 | 509.5 | 17.6 | 492 | G70899 | probable monooxyge |
| 7 | 500.5 | 17.2 | 495 | A70782 | probable monooxyge |
| 8 | 497.5 | 17.1 | 524 | D70861 | probable monooxygen |
| 9 | 466.5 | 16.1 | 527 | A83453 | probable monooxyge |
| 10 | 362.5 | 12.5 | 489 | C70655 | monooxygenase, fla |
| 11 | 359 | 12.4 | 499 | E87416 | probable monooxyge |
| 12 | 346.5 | 11.9 | 494 | A86917 | conserved hypothet |
| 13 | 342.5 | 11.8 | 499 | B83493 | arylesterase/monox |
| 14 | 339 | 11.7 | 833 | G75621 | monooxygenase, fla |
| 15 | 337 | 11.6 | 498 | G87567 | probable monooxygen |
| 16 | 326 | 11.2 | 486 | D70932 | hypothetical prote |
| 17 | 297 | 10.2 | 495 | G70852 | probable flavin-co |
| 18 | 232 | 8.0 | 431 | H84742 | monooxygenase, fla |
| 19 | 219 | 7.5 | 458 | C87345 | hypothetical prote |
| 20 | 216.5 | 7.5 | 533 | T08587 | dimethylaniline mo |
| 21 | 216.5 | 7.5 | 533 | S71617 | T29M8.12 protein - |
| 22 | 216 | 7.4 | 530 | B86326 | dimethylaniline mo |
| 23 | 213.5 | 7.4 | 533 | A46677 | hypothetical prote |
| 24 | 212 | 7.3 | 426 | T04527 | hypothetical prote |
| 25 | 212 | 7.3 | 528 | T22583 | hypothetical prote |
| 26 | 210 | 7.2 | 421 | T00955 | flavin-containing |
| 27 | 210 | 7.2 | 514 | A45912 | microsome flavin |
| 28 | 210 | 7.2 | 514 | A54250 | dimethylaniline mo |
| 29 | 207.5 | 7.2 | 535 | B35182 | |

| | | | | | | |
|----|-------|-----|-----|---|--------|--------------------|
| 30 | 206.5 | 7.1 | 346 | 2 | E70715 | hypothetical prote |
| 31 | 206.5 | 7.1 | 548 | 2 | T23436 | hypothetical prote |
| 32 | 202.5 | 7.0 | 437 | 2 | G86178 | hypothetical prote |
| 33 | 202.5 | 7.0 | 531 | 2 | S71618 | dimethylaniline mo |
| 34 | 200.5 | 6.9 | 531 | 2 | B54096 | flavin-containing |
| 35 | 200 | 6.9 | 469 | 2 | AG3139 | flavin-containing |
| 36 | 200 | 6.9 | 605 | 2 | E98148 | hypothetical prote |
| 37 | 199 | 6.9 | 458 | 2 | T37052 | probable flavin-co |
| 38 | 197.5 | 6.8 | 533 | 2 | S51131 | flavin-containing |
| 39 | 197 | 6.8 | 406 | 2 | T36632 | probable oxidoredu |
| 40 | 194.5 | 6.7 | 529 | 2 | T23431 | hypothetical prote |
| 41 | 192.5 | 6.6 | 558 | 2 | S29125 | dimethylaniline mo |
| 42 | 189.5 | 6.5 | 555 | 2 | A54096 | flavin-containing |
| 43 | 188.5 | 6.5 | 600 | 2 | G95854 | probable monooxyge |
| 44 | 187 | 6.4 | 415 | 2 | T07706 | hypothetical prote |
| 45 | 181 | 6.2 | 533 | 1 | S18380 | dimethylaniline mo |

ALIGNMENTS

RESULT 1
JC7158
steroid monooxygenase (EC 1.14.99.-) - Rhodococcus rhodochrous
C:Species: Rhodococcus rhodochrous
C:Date: 04-Mar-2000 #sequence_revision 04-Mar-2000 #text_change 21-Jul-2000
C:Accession: JC7158; PC7046
R:Moril, S.; Sawamoto, S.; Yamauchi, Y.; Miyamoto, M.; Iwami, M.; Itagaki, E.
J. Biochem. 126, 624-631, 1999
A:Title: Steroid monooxygenase of Rhodococcus rhodochrous: Sequencing of the genomic
A:Reference number: JC7158; MUID:99398443
A:Accession: JC7158
A:Molecule type: DNA
A:Residues: 1-549 <MOR>
A:Cross-references: DDBJ:AB010439; NID:2804297; PIDN:BAA24454.1; PID:g2804298
A:Experimental source: strain IFO 3338
A:Accession: PC7046
A:Molecule type: protein
A:Residues: 1-15;70-87;95-114;172-181;194-211;312-319;335-352;413-422;473-492;518-533
C:Comment: This enzyme is a Baeyer-Villigerase catalyzing the insertion of an oxygen
ble protein having catalytic properties.
C:Genetics:
A:Gene: smo
A:Start codon: TTG
C:Keywords: FAD; monooxygenase; nucleotide binding; oxidoreductase

| | | | | | | | | | |
|-----------------------|-------|---|---------|------------|--------|--------|---|------|---|
| Query Match | 30.9% | Score | 895.5 | DB 2 | Length | 549 | | | |
| Best Local Similarity | 36.3% | Pred. No. | 3.6e-62 | | | | | | |
| Matches | 192 | Conservative | 103 | Mismatches | 227 | Indels | 7 | Gaps | 6 |
| Qy | 16 | DAIIIVAGLSGSIVYKRLKRLNAKIFEGAPDPGGVHWNRYPGARVDSPTFYQLNI | - | 74 | | | | | |
| | | : : : : : : : : : : : | | | | | | | |
| Db | 23 | DVVVVGAGIAGLYAIHFRFSQGLTVRAFEAASGVGVWYNNRYPGARCDVSDISYFS | 82 | | | | | | |
| | | : : : : : : : : : : : | | | | | | | |
| Qy | 75 | PEVMDTWTSRCYRPOKELLSVYHCDKIRGLRKDYVFGAEVVDYARDLGTWTVTSA | 134 | | | | | | |
| | | : : : : : : : : : : : | | | | | | | |
| Db | 83 | PELEQENWSEKATQPELLAYLEHVADFRLDRDIRFTRVTSAVLDEGLRWTVTRDR | 142 | | | | | | |
| | | : : : : : : : : : : : | | | | | | | |
| Qy | 135 | GHVATAKYLLIATGLLHRKHPTALPLGLADPFGKVIHSSAW-HEDFDAEGORVAVIGAGAT | 193 | | | | | | |
| | | : : : : : : : : : : : | | | | | | | |
| Db | 143 | GDEVSAARFLVVAAGPLSNANTPAFDGLDRFTGDIVHTARWPHDGVDFTKRRVIGTSS | 202 | | | | | | |
| | | : : : : : : : : : : : | | | | | | | |
| Qy | 194 | STQIVOEAKKADQVTFMRRPFSYCLPMQRQRTMDRNEQTAKKAYYPTLFPAESKSRIGFP | 253 | | | | | | |
| | | : : : : : : : : : : : | | | | | | | |
| Db | 203 | GIQSPIITAEQAEQLFVFORANSISIPAGNVPLDDATRAEQKANYAERRLSRESGGSP | 262 | | | | | | |
| | | : : : : : : : : : : : | | | | | | | |
| Qy | 254 | VQAPSVGIFEVSPGEOREAYFEELWEGAFNFLACQYREVMDKKANRLVDFWAKKTRSR | 313 | | | | | | |
| | | : : : : : : : : : : : | | | | | | | |
| Db | 263 | HRHPKSALEVESEERRAYVEERKLGGLVLF-SKAPDQLTDPAAANDTARAFWEKIRAV | 321 | | | | | | |
| | | : : : : : : : : : : : | | | | | | | |
| Qy | 314 | IVNPAKRDLMAPLEPPYTFWFGTKRSPLESYYEMLDKPSVEIVNLEQSPISVAVTKTGVL | 373 | | | | | | |
| | | : : : : : : : : : : : | | | | | | | |

Db 322 VDDPAVAELLTPKD--HAIGAKRIVTDSGGYYETYNRDNVLDLURSTPIVGMDETGI-VT 378
QY 374 DGSKRECDTIVLATGFSFTGSLTHMGLKNKHGVDLKEVWKDGIISTYMGVSHGFPNAPF 433
Db 379 TGAHYDLDIMVLAFGDMATGSLDKLEIVGRGGRTLKTETWAGAPRTYILGLGIDGFPNPN 438
QY 434 VATAQAPTVLSNGPTIETQVDLIADTIKLEAEHATSVEATKSAQAEAWSIMIAKMNHT 493
Db 439 LTGPGSPSVLANMVLHSELHVDWADAIAVLDARGAAGIEGTPAIVADWVEECRNABAS 498
QY 494 LFPLTDSWMTGGNIPGRATRALTFIGGIALYEQICQEV-ANWDGFDVL 541
Db 499 LLANSANWYLGANLPGRPRVMPPLGGFGVYREITTEVAESGYGFAIL 547

RESULT 2
C87472
steroid monooxygenase [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
R:Wierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.B.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: C87472
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-556 <STO>
A:Cross-references: GB:AE005673; NID:g13423231; PIDN:AAK23775.1; GSPDB:GN00148
C:Genetics:
A:Gene: CG1799

Query Match 28.9%; Score 838; DB 2; Length 556;
Best Local Similarity 33.4%; Pred. No. 1.2e-57;
Matches 179; Conservative 104; Mismatches 235; Indels 18; Gaps 6;
QY 16 DAIIAGAGLSGISAIVKLRKLRLNAKIFEGAPDFGVMHNRYPGARVDSSTPFYQLNI- 74
Db 17 DALVIGAGGGVAVHKLSSGSLVSGFAGGDDGVWYNNRYPGARCDLMSLDYCYFS 76
QY 75 PEWKDWTWSCRPYDQKELLSYVHCDKIRGLRKDYVFGAEVVDARYDLGTWTKTSA 134
Db 77 PDIEQEWTSQPAQSEILAYANFVADRLDRHFLFTWVTGAAYDEANWVRVTVD 136
QY 135 GHVATAKYLILATGLLHRKHTPALPGLADPENGKVIHSSAW-HEDFDAEGORVAVIGAGAT 193
Db 137 GRVFATYCVMASGGLSPKGVFPDGDADDFKGBILAAKWPHTPVSPFEGKRVLVGTST 196
QY 194 SIQIVOLAKKADQVTFMRRPSYCLPMRQRTMDRNEQTAWKAYIPTLFESRSRIGFP 253
Db 197 GQIVPEVAKTAGALVVFQRTFSTPLPMRNVLTPEYVIELKKNYAGIRVARNSPJGV 256
QY 254 VOAPSVGIFEVSPQREAYFEELWEGAFNFLACQYREVMDKKNRVLVDYFMAKTRSR 313
Db 257 RQGSTRPFFSVTPQRRALMEDSWARGSLFLG-TFSDLLVNPPEANQVAFYRSKISEV 315
QY 314 IYNPAKRDLMAPLEPPYFVGTRKSPLESYEMLDKPSVEIVNLEOSPVIYVTKTGVL 373
Db 316 VTDPTVAEKLKRGYPIF--ARRPCLDTHYETFNLPVKLHDCLESPIKITPAG-LLT 372
QY 374 DGSKRECDTIVLATGFSFTGSLTHMGLKNKHGVDLKEVWKDGIISTYMGVSHGFPNAPF 433
Db 373 ETGVEVDLVIATGYDGLTGALLNPDVVGRLQDRKWRGAKSHLGLMIASFNNFL 432
QY 434 VATAQAPTVLSNGPTIETQVDLIADTIKLEAEHATSVEATKSAQAEAWSIMIAKMNHT 493
Db 433 VCGPGNPAALANIITLDQNDIWDIADAITHMQEGLATIEPTGAEQGWMDVYGMAL 492
QY 494 LFPLTDSWMTGGNIPGKATRALTFIGGIALYEQICQ-----EKVANWDG 537

Db 493 LVSKANTWVGGNISKPRGLSNYTGFGQRYSEACRLAAEKGYEDFVFERAATKQG 548

RESULT 3
A28550
cyclohexanone monooxygenase (EC 1.14.13.22) - Acinetobacter sp.
C:Species: Acinetobacter sp.
C:Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 08-Oct-1999
C:Accession: A28550; B28550
R:Chen, Y.C.J.; Peoples, O.P.; Walsh, C.T.
J. Bacteriol. 170, 781-789, 1988
A:Title: Acinetobacter cyclohexanone monooxygenase: gene cloning and sequence deter
A:Reference number: A28550; MUID:88115180
A:Contents: NCIB 9871
A:Accession: A28550
A:Molecule type: DNA
A:Residues: 1-542 <CHI>
A:Cross-references: GB:M19029; NID:g141767; PIDN:AAA21892.1; PID:g141768
A:Accession: B28550
A:Molecule type: protein
A:Residues: 2-11 <CH2>
C:Comment: The initiator Met is not shown.
C:Keywords: oxidoreductase

Query Match 27.6%; Score 802; DB 2; Length 542;
Best Local Similarity 33.4%; Pred. No. 7.6e-55;
Matches 179; Conservative 96; Mismatches 237; Indels 24; Gaps 8;
QY 16 DAIIAGAGLSGISAVYKLR-KLRNNAKIFEGAPDFGVMHNRYPGARVDSSTPFY---- 70
Db 7 DAIVGGGFGGLYAVKLRLELAKVQAFDKATDVAGTWYNNRYPGALLDTETHLYCYSW 66
QY 71 ---QLNIPEVWKDWTWSCRPYDQKELLSYVHCDKIRGLRKDYVFGAEVVDARYDLGT 137
Db 67 DKELQSLKIRK-----KYVGGPDVRKYLQOVAEKHDLKSYQFNATVQSAHYNEADAL 120
QY 128 WTVKTSAGHVATAKYLIILATGLLHRKHTPALPGLADPENGKVIHSSAWHEDFDAEGORVAV 187
Db 121 WEVTEYGDVKTARPLITAGLLSAPNLPNKGINQFQGLHHTSRWPDVDFEGKRGV 180
QY 188 IGAGATSIQIVOLEAKKADQVTFMRRPSYCLPMRQRTMDRNEQTAWKAYIPTLFESRK 247
Db 181 IGTGTVGVITAVAPLAKHLTVFQSAQSVPIGNDPLSEEDVKKIKDNYDKSLGCMWN 240
QY 248 SRIGFPVQAPSVGIFEVSPQREAYFEELWER-GAFNFLACQYREVMDKKNRVLVDYFW 306
Db 241 SALAFALNESTVPAKSVSAEERKAVFERAWQTGGGFRPFMTFGDIATNMEANIEAQNF 300
QY 307 AKTRSRIRVNPAPKRDLMAPLEPPYFVGTRKSPLESYEMLDKPSVEIVNLEOSPVIYV 366
Db 301 KGKIAEIVKDPAAIAQKLMQD-----LYAKRPLCDSSGYNTFNRDNVRLDEDVKANPVEIT 356
QY 367 KTGVLSDGSKRECDTIVLATGFSFTGSLTHMGLKNKHGVDLKEVWKDGIISTYMGVSH 426
Db 357 ENGKLENGDPVELDMLICATGFDVADGVNVRMDIQGNGLAMKDYWKEGPSSYMGVTYN 416
QY 427 GFNAPFVATAQAQPTVLSNGPTIETQVDLIADTIKLEAEHATSVEATKSAQAEAWSIM 486
Db 417 NYPNMFVLPNGNP--FTNLPSPSIESQVEWISDTQYTVENNVSIEATKEAEQWOTQC 474
QY 487 AKMNEHTLFPLTDSWMTGGNIPGKATRALTFIGGIALYEQICQEVAN--WGDV 540
Db 475 ANIAEMTLFPKQASWIFGANIPGKNTVYVYGLGLKEY-RTCASCKNHAYEGFDI 529

RESULT 4
E87474
steroid monooxygenase [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: E87474


```
QY 17 AIIVAGLSGISAVYKLRKRL--NAKIFEGAPDFGCVHWHNRYPGARVDSPTFFYQLNIP 75
Db 7 ALIIVAGFSGIGAAIKLDRAGFSYLVVEAGDVGVTWHNTYPGTAIVDSFYSQFSFE 66
QY 76 EVWKDWTWSCRYPQKELLSYVHCHDKIRGLRKDYVFGAEVVDARYARDLGTWTVTSAG 135
Db 67 Q---SRHWSRTYAPGHEUKAETHCVDKYGIRNSRLRNTKVLAEEFDDSHSLWRVQTDPG 123
QY 136 HVATAKYLILATGLLHRKHHTPALPLADFNKGKVIHSSAWHEDFDAEGORVAVIGAGATSI 195
Db 124 GEITAREFLISAGCILTVPKLPDIDGVDSFEGVTMHTARNDHTQDLTGKRVGIIGTASAV 193
QY 196 QIVQELAKKADQVTFMRRPSYCLPMRQRTWDRNEQTA--W-----KAYTPTFEASR 246
Db 184 QVIPEMAPIVSHLVFQRTPIWCFP-----KFDVPLPTAVRWAMRIPGGKAVHRLLSOAFV 239
QY 247 KSRIGRPVQAPSVGIFESPQREAYFEELMERGAFLACQYREVWVDKKANRLVYDWD 306
Db 240 EAT--FFIAHYFAVFLPAKHWSA-----G 263
QY 307 AKKTRSRIVPAKRDMLAPLEPPYWFGRKSPLESDDYEMLDKPSVEIVNLEQSPIVAVT 366
Db 264 RYLRQOVHDPVYRE---QLTPRAYGCKRPGFHTYLTSTFNRDNRLVT---EPIDKIT 317
QY 367 KTGVLSDGSKRECDTIVLATGDFDSF--TGSUTHMGLKNKHGVDLKEVWKD--GISTYMGYF 424
Db 318 PTAVATTDGASHIEDLVLATGPKRVLDTSIPTVAVTGTGASLSRFWDEHRLQAYEGVS 377
QY 425 SHGFPNAFFVATAQAPTVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQEAWSI 484
Db 378 VPGYPN--FTTVGPGYGVGSSYFALIIETQAHIIIRCLRARRTGATRIEVTBEANARYFA 436
QY 485 MIAKMNDEHTLF-----PLTDSW--TGNIPGKAT 512
Db 437 EVMRRRHQVFWQDSCRANSYYFDKNGDVLPRPT 471
RESULT 7
A70782
probable monooxygenase - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70782
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: A70782
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-495 <COL>
A:Cross-references: GB:273101; GB:AL123456; NID:g3261565; PIDN:CAA97398.1; PID:e241788;
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv0892
Query Match 17.2%; Score 500.5; DB 2; Length 495;
Best Local Similarity 28.8%; Pred. No. 3e-31;
Matches 143; Conservative 77; Mismatches 217; Indels 59; Gaps 12;
QY 15 CDAL-IVGAGLSGISAVYKLRKRL--NAKIFEGAPDFGCVHWHNRYPGARVDSPTFFYQL 72
Db 5 CPTAVVAGMCMCAVITLLSAGITDVCIEKADDDVGGTWRDNTYPGTLTCDVPSRLYQY 64
QY 73 NIPEVWKDWTWSCRYPDQKELLSYVHCHDKIRGLRKDYVFGAEVVDARYARDLGTWTVKT 132
Db 65 SPA---KNPNWTQMFSGGGIQQDYLGRGIAERYGURHIFGATVVSARF---DQGRWLRT 119
QY 133 SAGHVATAKYLILATGLLHRKHHTPALPLADFNKGKVIHSSAWHEDFDAEGORVAVIGAGA 192
```

```
Db 120 DSGTESTVDFLLISATGVLHHPRIPIAGLDDFRGTVPFHSARWHDHTVPLLGRRIAVIGTGS 179
QY 193 TSIQIIVQELAKKADQVTFMRRPSYCLPM---RQRTMDRNEQTAWKAYPTLFEASRKS 249
Db 180 TGVQLVCLGAGVAGVGTMTFQRTAQWVLPMPNPRYSKLARVFRHRAFPCLGSLAYKAYSLSF 239
QY 250 IGFPVQAPSVGIFESPQREAYFEELMERGAFLACQYREVWVDKKANRLVYDWDFAKK 309
Db 240 ETEVALSNPGL-----HRKLVCAVCRA 262
QY 310 TRSRIVPAKRDMLAPLEPPYWFGRKSPLESDDYEMLDKPSVEIVNLEQSPIVAVTGTG 369
Db 263 SLRRVRDPRLRALTPDYEPM---CKRLVMSGGFYRAIQRDDELVT---AGIDIVEHRG 316
QY 370 VLLSDGSKRECDTIVLATGDFDSFTGSLTHMGLKNKHGVDLKEVWKD--GISTYMGYFSGFP 429
Db 317 IVTDGVLHEVDVIVLATGDFDSH-AFFRPMQLTGRDGIIRDDWQDGPQHAHQTVAIQFPP 375
QY 430 NAFFVATAQAPTVLSNGP--TIETQVDLIADTIAKLEAEHATSVEATKSAQEAWSIMIA 487
Db 376 NFFMWLGPSP--VGNFPLTAVAESQAQEHIVQWIKMRHCEFDTPMEPKSAATEAYNTVLR 433
QY 488 KMNEHTLPL-TDSW 502
Db 434 AAMPNTVMTTGCDSWY 449
RESULT 8
D70861
probable monooxygenase - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: D70861
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: D70861
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-524 <COL>
A:Cross-references: GB:AL021287; GB:AL123456; NID:g3261508; PIDN:CAA16134.1; PID:e1
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV3049c
Query Match 17.1%; Score 497.5; DB 2; Length 524;
Best Local Similarity 28.7%; Pred. No. 5.6e-31;
Matches 142; Conservative 78; Mismatches 218; Indels 57; Gaps 13;
QY 17 AIIVAGLSGISAVYKLRKRLNAKIFEGAPDFGCVHWHNRYPGARVDSPTFFYQLNIP 76
Db 25 AVIIGTGFGLGMAIALQKQGVDFVILEKADDDVGGTWRDNTYPGCACDIPSHLYSFSF-E 83
QY 77 VWKDWTSCTPYDQKELLSYVHCHDKIRGLRKDYVFGAEVVDARYARDLGTWTVKTSAGH 136
Db 84 PKADWKHLFSYWD--EILGYLKGVTDKYGLRRIEFNSLVDRGYWDDDCRHFVFTADGR 141
QY 137 VATAKYLILATGLLHRKHHTPALPLADFNKGKVIHSSAWHEDFDAEGORVAVIGAGATSIQ 196
Db 142 EYVNOFLISGAGALHIPSFEIAGHDEFAGPAPFASQAWDHSIDLTKRVAIVGTGASAIQ 201
QY 197 IVOELAKKADQVTFMRRPSYCLPMRQRTMDRNEQTAWKAYPTLFEASRKSRIQFPVOA 256
Db 202 IVPDIVGQVAELQLYQRTPPWVPRTNBELPVLSRLALRT--VPLRALLR----- 250
QY 257 PSVGIFESPQREAYFEELMERGAFLACQYREVWVDKKANRL--VYDFWAKTKRSRIV 315
```

Db 251 --LGII-----WAEALAY-----GNTKRPNTLKIIEAYAKYNIRSV 286
QY 316 NPAKRDLMAPLEPPYWGTRKSPLESDYYEMLDKPSVEIVNLEOSPPIVAVTKTGVLSDG 375
Db 287 K--DRELRRKLTPRYICCKRIILSSYYPAVADPKTELIT---DRIDRITHDGIIVTADG 341
QY 376 SK----RECDFIVLATGFDSTGSLTHMGLNKKHGVDLKEVW-KDGISTYMGVFSHGFPN 430
Db 342 TGRVVFREADVIVATGF-HVTDSTYTVQIKRGHGEDLVDRNREGIGHRGITVANPNP 400
QY 431 AFFVATAQAPTVLSNGPTIETQVDLIADTIKLEAEHATSVEATKSAQEAWSMIAMKN 490
Db 401 LFFLLGPNLTGLHNSVVFMIESQIHVYADALAKCDRMGVQALATPRAQD-----RFN 453
QY 491 EHTLFPPLTDSWWTGG 505
Db 454 QELQRLLAGSVWNSG 468

RESULT 9
A83453
Probable flavin-containing monooxygenase PA1538 [imported] - Pseudomonas aeruginosa (str
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: A83453
R:Stover, C.K.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337
A:Accession: A83453
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-527 <STO>
A:Cross-references: GB:AE004582; GB:AE004091; NID:g9947492; PIDN:AAG04927.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1538

Query Match 16.1%; Score 466.5; DB 2; Length 527;
Best Local Similarity 25.2%; Pred. No. 1.5e-28;
Matches 127; Conservative 80; Mismatches 223; Indels 73; Gaps 10;
QY 15 CDALIVAGLSGISAVYKLRKLNKIFEGADPF-----GGVHNNRYPGARVDS 65
Db 26 CKVAIIIGTGSGLGMAIRLQ-----ECEDDFLIFEKADAGVGGTWRVNNYPGCACDV 77
QY 66 ETFFYQLNI---PEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLKRDYVFGAEVVDARYA 122
Db 78 QSHVYSFSEFANPE-----WTRFARQPEIRAYLEKCEWERYLQEKTLNTEIGKLAWD 131
QY 123 RDLGTWTVKTSAGHVATAKYLLIATGLLHRKHTPALPLGLADFNKGKVIHSSAWHEDFDAQE 182
Db 132 ERQSLMHLHDAQNHVYANAVVSGMGLSTPAYPRLDGLENFQCKVPHSQOWDHDYDLKG 191
QY 183 QRVAIVAGCATSIOIYQELAKKADQVTMFMRPSSYCLPMPQRTMDRNEQTAWKAAYPTLF 242
Db 192 KRVAIVGTGASATQFVPEIQPLVAALDYQRTPPWILPKPDRAISETERRFR-----245
QY 243 EASRKRIGFPVQAPSVGIFEVSPEQREAYFEELWERAFAENFLACQYREVWVDKKNRLV 302
Db 246 -----FPL-----VQLWRGGLYSLLEGRVLGTFAPQVWKLV 278
QY 303 YDFWAKKTRSRIVNPAKRDLMAPLEPPYWGTRKSPLESYIEMLDKPSVEIVNLEQSPI 362
Db 279 QRLAIRHIHKIKDP---ELRRKVTDPDTIGCKRIILSMHNYYPALAAANSTVIT---EGI 332
QY 363 VATKTKGVLLSDGSKRECDTIVLATGFDSTGSLTHMGLNKKHGVDLKEVWKGDIISTYMG 422
Db 333 RAVTANGIVDNGRERREVDALIFCTGFTA-NDPIPRGVVFPGRDGRDLSDWSKGPAYKG 391

QY 423 VFSHFPPNAFFVATAQAPTVLSNGPTIETQVDLIADTIKLEAEHATSVEATKSAQEA 482
Db 392 TTTAGFPNLFLLGMPNTGLGHNSVWYMIESQIAYVLDALKMKRRRELLSLEKAPVGE-- 449
QY 483 SIMIAKMNSEHTLFPPLTDSWWTGG 505
Db 450 -----RYNEYLORKLDRSVWSVG 467
RESULT 10
C70655
Probable monooxygenase - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: C70655
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete ge
A:Reference number: A70500; MUID:98295987
A:Accession: C70655
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-489 <COL>
A:Cross-references: GB:Z83864; GB:AL123456; NID:g3261687; PIDN:CAB06212.1; PID:e301
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv3854C

Query Match 12.5%; Score 362.5; DB 2; Length 489;
Best Local Similarity 26.3%; Pred. No. 2e-20;
Matches 124; Conservative 66; Mismatches 187; Indels 95; Gaps 18;
QY 16 DALIVAGLSGISAVYKLRKLNKAK---IFEGAPDFGGVHNNRYPGARVDSPTFYQL 72
Db 6 DVIVGAGISGVSAAHQLQD-RCPTKSYAILEKRESMGGTWDLFRYPGRSDSD--MYTL 62
QY 73 NIPEWKDWTWSCRYPDQKELLSYVHHCDKIRGLKRDYVFGAEVVDARYDLGTWTVK- 131
Db 63 GFR--FRPWTGQATADGKPILEYYKYSTAAMYGDIDRHFHKVISADHSTAENRTVHI 120
QY 132 TSAGHVA--TAKYLILATGLLH--RKHTPALPLGLADFNKGKVIHSSAWHEDFDAQE 187
Db 121 QSHGTLALTCEFLCISYNNYDEGYSRFAESDFVGPPIIHPQHWPELDLYDAKNIV 180
QY 188 IGAGATSIQIYQELAKK-ADQVTMFMRPSSYCLPMPQRTMDRNEQTAWKAAYPTLF 246
Db 181 IGSATAVTLVPALADSGAKHVTMLQSPYIVSQPDRDGIAEKLRW-----228
QY 247 KSRIGFPVQAPSVGIFEVSPEQREAYFEELWE---RGAENFLACQYREVWVDKKNRLV 303
Db 229 -----LPETM-AYTAVRWKNVLQAAVYSACQK-----255
QY 304 DFWAKKTRSRIVNPAKRDLMAPLEPPYWGTRKSPLE-----SDYIEMLDKPSVEIV 355
Db 256 --WPRMRKMFLLSIQRLQPEGYDVRKHFGPHYNPDQRLCLVPNGDGLFRATRHHGKVEV 313
QY 356 NLEQSPIVAVNKTGVLLSDGSKRECDTIVLATGFD--SFTGSLTHMGLNKKHGVDLKEVW 413
Db 314 T---DTIERFATGIRLNSGRRLPADIIITATGLNLQLEFGGATA-----TIDGQOVD 362
QY 414 KDGISTYMGVFSHGFPN-AFFVATAQAPTVLSNGPTIETQVDLIADTIADIAKL 464
Db 363 ITTMAYKGMMLSGIPNNMAYTVGYTNASWTL-----KADLYSEFCRL 405
RESULT 11
E87416
monooxygenase, flavin-binding family [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus

A;Molecule type: DNA
A;Residues: 1-494 <STO>
A;Cross-references: GB:AL450380; NID:g13092457; PIDN:CAC29573.1; GSPDB:GNO0147
C;Genetics:
A;Gene: ML0065

| | Query Match | 11.9% | Score 346.5; | DB 2; | Length 494; |
|----|-----------------------|---|--------------------|------------|-------------|
| | Best Local Similarity | 25.6%; | Pred. No. 3.6e-19; | | |
| | Matches 121; | Conservative 70; | Mismatches 185; | Indels 97; | Gaps 19; |
| Qy | 16 | DAIIVGAGISGSIVYKLRKLRLNAK--IFEGAPDFGGVWHNNRYPGARVDSETPFYQLN | 73 | | |
| | : : : | : | : | : | |
| Dd | 11 | DWIVVGAGISGVSAAWHLQDRCNKSVILEKRACMGGTWDLFRYPGLRSDSD-NYTLC | 68 | | |

69 FR--FRPWTKQALADGKPILEYIKSTAVMHGIDKIRLHNHKVTGADWSSIENRTWTVQVE 126 Db

132 -TSAGHVATAKYLLILATGLLH--RKHTPALPLGLADFNKGKVIHSSAWHEDFDAECQVAVI 188 Qv

Db 127 NNGTPRMISCSFLCSCGYNVEQGYPTFLGSEDFTGPIIHPQHWPEDLDYAAKNIVVI 186

QY 189 GAGATSIQIVQELAKK-ADQVTMFMRPPSYCL--PMRQRTMDRNEQTAWKAYPTLFEAS 245

187 CSCATAVTIVPAIANSGAKHVMTLORSPTYIVSOPAKDKI-----AA 228

[illegible]

QY 246 RKSRI GFFVQAPSVGIFEVSPQE RAYFEELWE ---RGAFNF LACQYKEVMVDKANKRLV 302

Db 229 RLNR-WLP-----DKYAYTAVRWKNILRQSALYGACQK----- 260

| | | | |
|----|-----|---|-----|
| QY | 303 | YDFWAKKTRSRIVNPAKROLMAPLEPPYWFCTKRSPLE-----SDYYEMLDKPSVEI | 354 |
|----|-----|---|-----|

```

261  ---WROBMBNIIICUVAAPBI  DKCYNOVKHECDBHNPWEORICLVPOGDIFRAIRKGTVDM  317

```

[illegible]

Qy 355 VNLEQSPIAVTKTGVLSDGSKRECDTIVLATGFD--SFTGSLTHMGLKNRKHGVDLREV 412

Db 318 VT---DAIDRFTSTGIRLKSGNELRADIIVTATGLNLQLFGGAVATV---DGQPVDLAQ 371

QV 413 WKDGISTYMGVFSHGFPNAFF-VATAQAPTVLSNGPTIIETQVDLIADTIACL 464

```

370      :| | : | | : | | : | | :: |
371      -----MEVYKMMI SCI DNMTVTVCYTNASWTL-----KADLVSEFFCRL 410

```

DD FORM 372 1-55 (REV. 11-15-53) GPO

RESULT 13

B83493

C:Species: *Pseudomonas aeruginosa*
conserved hypothetical protein PA1208 [Imported] - *Pseudomonas aeruginosa* (strain ATCC 27803)

```
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: 82402
```

C/Accession: B83493
R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey,

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbi9,
i. Forv s. Olson M V

Dr. LLOYD, S. C. Wilson, M.V.
Nature 406, 959-964, 2000

A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PAU1, an opportunistic pathogen
A:Reference number: A829950; MUID:20437337

A;Accession: B83493

A;Status: preliminary
A;Molecule type: DNA

A;Residues: 1-499 <STO>

A; Cross-References: GB:AE004531, GB:AE004691, NID:9557133, (E)NID:9557133, (E)
A; Experimental source: strain PAO1

C:Genetics:
A:Gene: PA1208

CONCLUSIONS

Query Match 11.8%; Score 342.5; DB 2; Length 499;

| Best Local Similarity | 26.0%; | Pred. No. 7.5e-19; | | |
|-----------------------|--------|--------------------|---------|------|
| Best Local Similarity | 26.0%; | Pred. No. 7.5e-19; | Indels | 89. |
| | | | Matches | 193. |
| | | | Indels | 13. |

| | | | | | |
|---------|------|--------------|----------------|--------------|----------|
| Matches | 114; | Conservative | 33; Mismatches | 103; Leaders | 95; Caps |
| | | | | | 76 |

Qy 16 DAIIVGAGLSGISAVYKLRKRLNAKIF--EGAPDFGGVHWNRYPGARVDSPTFYQL 72

Db 7 DVLIVGAGLSGVGAAYHLMK-HCPGKSFALLEGRAAMGTTWDLFRYPGIRSDSD---MF 61


```
Db 257 -----PEVKERLLGMV-RDHLGPGYDVETHFTPRYNPDORLCLVPDADLFDAIKSGAA 310
Qy 353 EIVNLEQSPIVAVTKTGVLISDGSKRCDTIVLATGPDSTGSLTHMGLKNKHGVDLKEV 412
Db 311 SVVT---DHIEFTTETGIQLKSGKTLADADVIVTATG-----LQLQLSGM---EV 354
Qy 413 WKDG-----ISTYMGVFESHGPPN 430
Db 355 VYDGKVADLSQSMSYKGMFSDVPN 379
```

Search completed: April 27, 2002, 07:48:20
Job time: 6727 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:46:19 ; Search time 24.91 Seconds
(without alignments)
153.038 Million cell updates/sec

Title: US-09-882-694A-3
Sequence: 1 MSATNSRGDCSVACDAIIV.....QICQEKVANWDGFDVLHAPC 545

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 40831 seqs, 6994805 residues

Total number of hits satisfying chosen parameters: 40831

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA.New:*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 216.5 | 7.5 | 414 | 5 | US-09-573-655B-2026 |
| 2 | 125.5 | 4.3 | 717 | 1 | PCT-US02-09921-814 |
| 3 | 106 | 3.7 | 320 | 5 | US-09-540-209B-8514 |
| 4 | 105.5 | 3.6 | 491 | 5 | US-09-540-209B-10395 |
| 5 | 97.5 | 3.4 | 5183 | 6 | US-10-107-521-1 |
| 6 | 92 | 3.2 | 486 | 5 | US-09-573-655B-279 |
| 7 | 85 | 2.9 | 539 | 6 | US-10-106-698-6386 |
| 8 | 83.5 | 2.9 | 1156 | 5 | US-09-573-655B-1983 |
| 9 | 83 | 2.9 | 526 | 5 | US-09-540-209B-9188 |
| 10 | 83 | 2.9 | 2171 | 5 | US-09-573-655B-291 |
| 11 | 82.5 | 2.8 | 844 | 5 | US-09-573-655B-267 |
| 12 | 82 | 2.8 | 565 | 6 | US-10-102-806-557 |
| 13 | 81 | 2.8 | 536 | 5 | US-09-540-209B-6020 |
| 14 | 81 | 2.8 | 947 | 5 | US-09-573-655B-117 |
| 15 | 80.5 | 2.8 | 327 | 5 | US-09-540-209B-6436 |
| 16 | 80.5 | 2.8 | 478 | 5 | US-09-540-209B-6630 |
| 17 | 79.5 | 2.7 | 703 | 5 | US-09-573-655B-2365 |
| 18 | 79.5 | 2.7 | 1251 | 5 | US-09-540-209B-7929 |
| 19 | 79 | 2.7 | 433 | 5 | US-09-540-209B-7557 |
| 20 | 79 | 2.7 | 652 | 5 | US-09-540-209B-7153 |
| 21 | 78 | 2.7 | 1167 | 5 | US-09-540-209B-9828 |
| 22 | 77.5 | 2.7 | 536 | 5 | US-09-573-655B-1461 |
| 23 | 77.5 | 2.7 | 1152 | 5 | US-09-540-209B-10027 |
| 24 | 77 | 2.7 | 456 | 5 | US-09-540-209B-9647 |
| 25 | 77 | 2.7 | 492 | 5 | US-09-762-154-10 |
| 26 | 77 | 2.7 | 1084 | 5 | US-09-540-209B-5327 |

| | | | | | | |
|----|------|-----|------|---|---------------------|-------------------|
| 27 | 76.5 | 2.6 | 408 | 6 | US-10-106-698-4724 | Sequence 4724, Ap |
| 28 | 76.5 | 2.6 | 446 | 5 | US-09-540-209B-7273 | Sequence 7273, Ap |
| 29 | 76.5 | 2.6 | 799 | 5 | US-09-540-209B-6802 | Sequence 6802, Ap |
| 30 | 76.5 | 2.6 | 1008 | 5 | US-09-540-209B-8074 | Sequence 8074, Ap |
| 31 | 76 | 2.6 | 255 | 5 | US-09-573-655B-199 | Sequence 199, App |
| 32 | 76 | 2.6 | 255 | 5 | US-09-573-655B-261 | Sequence 261, App |
| 33 | 76 | 2.6 | 255 | 5 | US-09-573-655B-1523 | Sequence 1523, Ap |
| 34 | 76 | 2.6 | 255 | 5 | US-09-573-655B-1753 | Sequence 1753, Ap |
| 35 | 76 | 2.6 | 255 | 5 | US-09-573-655B-1867 | Sequence 1867, Ap |
| 36 | 76 | 2.6 | 450 | 6 | US-10-105-299-3365 | Sequence 3365, Ap |
| 37 | 76 | 2.6 | 450 | 6 | US-10-105-299-3367 | Sequence 3367, Ap |
| 38 | 76 | 2.6 | 533 | 6 | US-10-108-605-193 | Sequence 193, App |
| 39 | 76 | 2.6 | 737 | 5 | US-09-540-209B-9281 | Sequence 9281, Ap |
| 40 | 76 | 2.6 | 763 | 5 | US-09-540-209B-8497 | Sequence 8497, Ap |
| 41 | 76 | 2.6 | 1465 | 1 | PCT-US02-07826-219 | Sequence 219, App |
| 42 | 76 | 2.6 | 1465 | 6 | US-10-097-340-219 | Sequence 219, App |
| 43 | 75.5 | 2.6 | 204 | 6 | US-10-102-806-743 | Sequence 743, App |
| 44 | 75.5 | 2.6 | 663 | 6 | US-10-108-605-245 | Sequence 245, App |
| 45 | 75.5 | 2.6 | 1572 | 5 | US-09-573-655B-1491 | Sequence 1491, Ap |

ALIGNMENTS

RESULT 1

US-09-573-655B-2026

; Sequence 2026, Application US/09573655B

; GENERAL INFORMATION:

; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim

; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypept

; TITLE OF INVENTION: Thereby

; FILE REFERENCE: 2750-0876P

; CURRENT APPLICATION NUMBER: US/09/573,655B

; CURRENT FILING DATE: 2000-05-18

; NUMBER OF SEQ ID NOS: 3281

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 2026

; LENGTH: 414

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

US-09-573-655B-2026

Query Match 7.5%; Score 216.5; DB 5; Length 414;

Best Local Similarity 21.9%; Pred. No. 1.5e-11;

Matches 97; Conservative 58; Mismatches 179; Indels 109; Gaps 14;

| | | | |
|----|-----|---|--------|
| Qy | 18 | IIIVAGLSGTSAVYKLRKLRNAKIFEGAPDFGCVHWNRYPGARVDSETPFYQ--- | LNI 74 |
| Db | 22 | IIIGAGPSGLATSACLSRGVPSLILERSDSIASLMSKSTYDRLLRLHLPKHFCKRLPLD | 81 |
| Qy | 75 | PEVWKDWTSRCRYPDQKELL-----SYVHCODKIRGLRKQVYFGAEVVDARYARDLGTWT | 130 |
| Db | 82 | PEYVP-----KYPKNEFLAYLESYASH-----FRIAPRFKNKNVQAAAYDSSSGFWRV | 129 |
| Qy | 131 | KTSAGHVATAKYLILATGLLHRKHTPALCLADFN-GKVIHSSAWHEDFDAQORVAVIG | 189 |
| Db | 130 | KTHDNTTEYLSKWLIVATGENADYPFPEIPGRKKFSGGKIVHASEYKSGEERFRQKVLVVG | 189 |
| Qy | 190 | AGATSIQIYOELAKKADQVTMFMRPSYCLPMRQRTMDRNEQTAWKAYTYPTLFEASRKS | 249 |
| Db | 190 | CNSGMEISLSDLVRRHNASHLVVNTVHVLP-----GNTLLKC-----LPLRLVDKFLLLM | 251 |
| Qy | 250 | IGFPVQAPSGVGFVSPEOREAYFEELWERGAFNFIACQYREVWVKANRLVYDFWAKK | 309 |
| Db | 222 | -----EILGVSTFGV-----GNTLLKC-----LPLRLVDKFLLLM | 251 |
| Qy | 310 | TRSIRVNPAKRDLMALEPPYWFCTKRSPL-----SDYYEMLDKPSVEIVNLEQSPIV-- | 363 |
| Db | 252 | ANLSFGNTDRLGLRRP-----KTGPLELKNVTGKSPVLDVGANSLRSGMIQIMEG | 302 |
| Qy | 364 | --AVTKTGVLSDGSKRECDTIVLATGFDSETGS-----LTHMGLKNKHGVDLKEVWK | 414 |

Db 303 VKEITKKGAKPMDQOEKDFDSIIIFATGYKSNVPTWLOGGDFTTDDGMPK---TPFPNGWR 359

QY 415 DGISTYMGVFSHGFPNFAFFVATA 437

Db 360 GG-----KGLYTVGFTTRGCLLCTA 378

RESULT 2

PCT-US02-09921-814

; Sequence 814, Application PC/TUS0209921

; GENERAL INFORMATION:

; APPLICANT: INCYTE GENOMICS, INC.

; APPLICANT: DAFFO, Abel

; APPLICANT: JONES, Anissa L.

; APPLICANT: TRAN, Alanna-Phung B.

; APPLICANT: DAHL, Christopher R.

; APPLICANT: GIETZEN, Darryl

; APPLICANT: CHINN, Joyce

; APPLICANT: DUFOUR, Gerard E.

; APPLICANT: HILLMAN, Jennifer L.

; APPLICANT: YU, Jimmy Y.

; APPLICANT: TUASON, Olivia

; APPLICANT: YAP, Pierre E.

; APPLICANT: AMSHEY, Stefan R.

; APPLICANT: DAUGHERTY, Sean C.

; APPLICANT: DAM, Tam C.

; APPLICANT: LIU, Tommy F.

; APPLICANT: NGUYEN, Duy-Viet An

; APPLICANT: KLEEFELD, Yael

; APPLICANT: GERSTIN JR., Edward H.

; APPLICANT: PERALTA, Careyna H.

; APPLICANT: DAVID, Marie H.

; APPLICANT: LEWIS, Samantha A.

; APPLICANT: CHEN, Alice J.

; APPLICANT: PANZER, Scott R.

; APPLICANT: HARRIS, Bernard

; APPLICANT: FLORES, Vincent

; APPLICANT: MARWAHA, Rakesh

; APPLICANT: LO, Audrey

; APPLICANT: LAN, Ruth Y.

; APPLICANT: URASHKA, Michael

; TITLE OF INVENTION: SECRETORY MOLECULES

; FILE REFERENCE: PT-1232 PCT

; CURRENT APPLICATION NUMBER: PCT/US02/09921

; PRIOR FILING DATE: 2002-03-27

; PRIOR APPLICATION NUMBER: 60/280,067; 60/280,068; 60/291,280; 60/291,849;

; 60/291,829; 60/299,428; 60/300,001; 60/299,776

; PRIOR FILING DATE: 2001-03-29; 2001-03-29; 2001-05-16; 2001-05-17;

; 2001-05-17; 2001-06-19; 2001-06-20; 2001-06-20

; NUMBER OF SEQ ID NOS: 1146

; SOFTWARE: PERL Program

; SEQ ID NO 814

; LENGTH: 717

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc.feature

; OTHER INFORMATION: Incyte ID No: LI:1018424.4.orf2.2001MAY17

PCT-US02-09921-814

Query Match 4.3%; Score 125.5; DB 1; Length 717;

Best Local Similarity 23.7%; Pred. No. 0.0025;

Matches 59; Conservative 36; Mismatches 103; Indels 51; Gaps 12;

QY 16 DAIIIVAGLSGISAVYKLRKLRNAKIFEGAPDFGGVWHNNRYPGAR-VDSSTPPY-----70

Db 65 DYCVLGAGPAGLQAYELQIRAGRDYAVFERAPRGS---FFTRYPRHRKLISINKRYTGKA 122

QY 71 --OLNIPVWKDWT-----WSCRY-PDQKELLSYVHCDKIRGLRKDYFGAEVV 117

Db 123 NAEFNLRHDWNSLLSHDPRLLFRHYSRAYFPDARDMVRYLGDFADTLGLR--VQYNTTIA 180

QY 118 DARYARDLGTWT-----VKTSGAHVATAKYLIILATGLLLHRKHTPALPCLADFNKQKVI--- 169

Db 181 HVTLDKDRQAWNGHYFILTQKQGVHQCISLVFATGL-----SVPNQVDFPGSEYAE 233

QY 170 HSSAWHEDFDARQORVAVIGAGATSIQIVQELAKKADQVTMFMRPSPCYCLPMRORTMDRN 229

Db 234 YESVSDPEDFVGQNVLIILGRNSAF-----ETAENILGVTFNFIHMLS-----RSRV----- 280

QY 230 EOTAWKAYY 238

Db 281 -RLSWATHY 288

RESULT 3

US-09-540-209B-8514

; Sequence 8514, Application US/09540209B

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDE

; FILE REFERENCE: 2709.1001-001

; CURRENT APPLICATION NUMBER: US/09/540,209B

; CURRENT FILING DATE: 2000-04-04

; NUMBER OF SEQ ID NOS: 10444

; SEQ ID NO 8514

; LENGTH: 320

; TYPE: PRT

; ORGANISM: B. fragilis

US-09-540-209B-8514

Query Match 3.7%; Score 106; DB 5; Length 320;

Best Local Similarity 24.7%; Pred. No. 0.04;

Matches 61; Conservative 33; Mismatches 79; Indels 74; Gaps 12;

QY 18 IIVGAGLSGISAVYKLRKLRNAKIFEGAPDFGGVWHNNRYPGARVDSET-----PFYQL 72

Db 14 LIIGSGPAGYTAAYAGRANLSPVLYEGI-----OPGQLTTTDDVENFPGY 62

QY 73 NI--PEVWKDWTWSCRYPDQKELLSYVHCDKIRGLRKOV-YFGAEVVVDARY----ARDL 125

Db 63 GISGPQLMED-----LRTQAEFRGA---DIRFGIATASDL 94

QY 126 G--TWTVKTSAGHVATAKYLIILATGLLLHRKHTPALPCLAD---FNGKVIHSSAWHEDFDA 180

Db 95 GOAPYKTIIDGKRVTEADSLIATGA-----TAKYILGLDDEKKYAGVGSACATCDGFFY 149

QY 181 EGQRVAVIGAGATSIQIVQELAKKADQVTMFMRP-----SYCLPMRORTMDRNEQTAWKA 236

Db 150 RKKVVAVVGGGTACEEATYLAGASKVYLVVRKPYLRASKIMQERVRRKDKIE----- 203

QY 237 YYPTLFE 243

Db 204 ---VLFE 207

RESULT 4

US-09-540-209B-10395

; Sequence 10395, Application US/09540209B

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDE

; FILE REFERENCE: 2709.1001-001

; CURRENT APPLICATION NUMBER: US/09/540,209B

; CURRENT FILING DATE: 2000-04-04

; NUMBER OF SEQ ID NOS: 10444

; SEQ ID NO 10395

; LENGTH: 491

; TYPE: PRT

; ORGANISM: B. fragilis

; FEATURE:

; NAME/KEY: UNSURE

| | | | | | |
|----|-----|--|--|--------------------------|------|
| Qy | 104 | RGLRKDVYFGAEVVD | -ARYARDLGTWTVTSAGH- | -----VATAKYLI-LATGILLHRK | 1530 |
| | | :I | :I | :I | |
| Db | 432 | KGKEKDPALAIRVDII | LSRTKEGVGPKLGQHGQGVLSILANHAIIKLTSVQFDL | 4910 | |
| | | :I | :I | :I | |
| Qy | 154 | HTPALPLGLADPENGKVIHSSAWHEDFDPAEGORVAV- | IGAGATSIQIIVQELAKKADQVTMEM | 2122 | |
| | | :I | :I | :I | |
| Db | 492 | QVEAL----- | HKQWETDGPAAALSIWAOSTSTORIORLDSVPLNNLL | 5350 | |

| | | | | |
|-----------------------|-------|----------------|-------|----------------------------------|
| Query Match | 3.2% | Score 92; | DB 5; | Length 486; |
| Best Local Similarity | 20.4% | Pred. No. 1.2; | | |
| Matches | 89; | Conservative | 56; | Mismatches 163; Indels 128; Gaps |

| | | | | | | |
|----|-----|----------------|------------------|------------------|-------------|------------------|
| Qy | 17 | AIIVG---- | AGLSGISAVYKLRKLR | LNAKIFEGAPD | GGVHHNRP | PGARVDSETP-FY 70 |
| | | : : | | : | | |
| Db | 137 | AVLLGDQKAVGGSG | -----KVIA | SKPNDFIVYADHGG | -----PGVLGM | ENTPHY 184 |
| Qy | 71 | QLNPTEWKWNTW | SCRDPDQKELLSYV | HHCDKIRGLRKD | VDYFGEAVDV | ARYARDLGTWV 130 |
| | | | | | | |
| Db | 185 | AADFTEITLKKH | ASCTY-----KEM | VIYVEACE-----SGS | IFEGIMPKD | LNIVV- 225 |

```

; SEQ ID NO: 279
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-279

```

```

Db      367  KKDTTLKELTETTRHRKHLQASVELIATI-----LFG-----PTMVLNL 406
Qy      358  EQSPIVAVTGTLLSDGSKRECDTIVLATFDSFGTSLTHMGLKN-----KHGVDL 409
Db      407  VREP-----GLPLVD--DWECLKSMVRV-FEEHCGSLTQYGMKHMRAFANVCNNGVS- 455
Qy      410  KEVWKDGISTYMGVFS 425
Db      456  KELMEEASTAACGGYS 471

```

```

RESULT 7
US-10-106-698-6386
; Sequence 6386, Application US/10106698
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005PI
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6386
; LENGTH: 539
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (2)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (7)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (11)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; US-10-106-698-6386

```

```

Query Match      2.9%; Score 85; DB 6; Length 539;
Best Local Similarity 19.7%; Pred. No. 5.5;
Matches 102; Conservative 49; Mismatches 175; Indels 192; Gaps 23;

QY    67 TPFVQLNIPWKKDWTWSCRYPDGKELLSVVHHCKIRGLRKDVYFGAEVVDARYADLG 126
      || : | : | : ||| : : || : | :
Db    14 TPSP--NTTHVRESW-YACRYRSGIPGSTHASATTVPGLSEE----- 53

QY    127 TWTVKTSAGHVATAKYLLILATGLLRHKHTPALPLGLADFNCVKI-----HSSAWH 175
      || : | : | : ||| : : || : | :
Db    54 STTFYSPGGSTETAFSHSNTMHSIQOSTFPDPSPGFTHVLPAITLTTDIGQUESTAFH 113

QY    176 EDFDAEQRVAVIGAGATSQIVQELAKKADQVTMRKRPSCYCLPMRQRTMDRNEQTAWK 235
      |||| : | : | : ||| : | : |||
Db    114 SSSDATG--YTPLPARSTASDLVE-----PTTFVISPS----- 145

QY    236 AYTPTLFEARSKSRIGFPPVQAPSYGIFEVSPEQREAYFEELWERGAFLNACLQYREVWVD 295
      |||| ||| : | : : ||| : | : |||
Db    146 PTVTTLFPAS-SSTSGLTEESTT---FHTSP----- 172

QY    296 KKANRLVYDFWAKTRSRIVNPAAKRDLMAP---LEPPYWFGTK-----RSPLE 340
      |||| : | : | : ||| : | : |||
Db    173 -----SFTSTIVISTESLETPLGCOEGOIWNGKCVCQQGVGYOCLSPLE 219

```

[illegible]

```

RESULT      8
US-09-573-655B-1983
; Sequence 1983, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVLEV, Victor and TROUKHAN,
; TITLE OF INVENTION: Sequence-Determined DN
; TITLE OF INVENTION:
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573,655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1983
; LENGTH: 1156
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-1983

```

| | | | | |
|-----------------------|--------|--|-----------------|----------------------|
| Query Match | 2.98; | Score 83.5; | DB 5; | Length 1156; |
| Best Local Similarity | 20.18; | Pred. No. 22; | | |
| Matches | 102; | Conservative 66; | Mismatches 175; | Indels 165; Gaps 24; |
| QY | 123 | RD LGTWTVK--TSAGHVATAKYL-----ILATGLLHRKH---PALPCLADFNCKV | 168 | |
| Db | 19 | RD FGN YGLKLESTWSSVSGRIKFEGCLVLGGIITWRGVLKERVVAISPATYMSQGN--- | 75 | |
| QY | 169 | IHS SAWHEDFPAEQRVAVIGAGATSIQI VOELA-----KKADQVTMF | 211 | |
| Db | 76 | --NSPRGDKDVE-----AGTSSSTEYEDSPDIASTKNAPVERLRNRQAALVLNA | 125 | |
| QY | 212 | MRRPSYCLPMRQ-----RTMDNRQETAWKAYYYITLFE--ASRKSRIGFPVQAPS YGI | 261 | |
| Db | 126 | SRRPRYTLDLKRREDDKQMLKRMRAHAQA IRAAH--LFKAASRVTGIASPLPTPGGSD | 182 | |
| QY | 262 | FEVSPQO-----REAYFEELWERCFAFNLACQYR-----EYWDVKANRLVYDF | 305 | |
| Db | 183 | FGIQEQIVSIRSDQNTGALQELGVRGLSDLLKTNLEKTHGDDDDTLKRKSAFGSN TY | 242 | |
| QY | 306 | WAKKTR S--RTVNP AKRDL-----MAPLEPPYFGTKRSPLESDDYEWLKPSPVEI | 354 | |
| Db | 243 | PQKGRSFWRPVWEASODLTLLIILVAAVSLA-----LGIKTEGIEKGWYDGIS----- | 292 | |
| QY | 355 | VNLEQSPIVAVTKGVLISDGSKRECDTIVLATGDFSFTGSLTHMGL--KNKHGVDLK E W | 413 | |
| Db | 293 | -----IAPAVLL-----VIVVTATSDYROSLOFQNLNEEKNRNL--EVT | 330 | |
| QY | 414 | KDG-----ISTY-----MGVFSHGFPNAPFVATAQAPTVLSNGPTIIET | 452 | |
| Db | 331 | RDGRVRVEISYDIYVGVDPVPLNIGDQASFMILSMQYFLALFRISVPADGVLVAGHS LAVD | 390 | |
| QY | 453 | QVDLIADT--IAKLEAEHATSV EATKSA-----QEAWSIMIAKNNEHTL PPL | 497 | |
| Db | 391 | ESSMTGSEKIVQKNSTKHPTLMSGCKVADGNGTMLVTGVGVNTWEMGLMLASVSD----- | 445 | |
| QY | 498 | TDSWMTGTGNIP--GKATRALTFIGGIAL | 523 | |

```
Db 446 -----NGGETPQVRLNGVATFIGTGL 468
                                     ||| | : |
US-09-540-2098-9188
; Sequence 9188, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FR
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709,1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 9188
; LENGTH: 526
; TYPE: PRT
; ORGANISM: B. fragilis
US-09-540-2098-9188

Query Match 2.9%; Score 83; DB 5; Length 526;
Best Local Similarity 25.1%; Pred. No. 8;
Matches 51; Conservative 33; Mismatches 75; Indels 44; Gaps 11;

QY 16 DAIVAGAGLSG-ISAVYKLRKLRNKAIFEGAPDFGGVWHNNRYPGARVDSETPFYOL-N 73
| | | | : | | | : | | | : | | | : |
Db 223 DVIVAGGPGAGSAAATYARK-GLNVAV1--AERIGG-----QVKETVGIENLIS 269
| | | | : | | | : | | | : | | | : |

QY 74 IPEWKDW-----TWSCRYPDQKELLSYVHCDKIRGLKDYFGEAEVVDARYARDLGT 127
| | | | : | | | : | | | : | | | : |
Db 270 VPSTTSQLADNLKTHMSQYP--IDLLEH-RQEKIEBLDGD-----KVLTA 314
| | | | : | | | : | | | : | | | : |

QY 128 WTVKTSAGHVATAKYLILATGLLRKHTPALPGLADPENGKVIHSSAWHEDFDAQGVAV 187
| | | | : | | | : | | | : | | | : |
Db 315 -----GGERFSAPIVATGASWRKLN--VPGAEVIGRVAFCPHCDGDFYKGGQVAV 366
| | | | : | | | : | | | : | | | : |

QY 188 IGAGATSIQIVQELAKKADQVTM 210
| | | | : | | | : | | | : |
Db 367 VGGNGSGIEAAIDLAGICSKVTV 389
| | | | : | | | : | | | : |

RESULT 10
US-09-573-655B-291
; Sequence 291, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573,655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 291
; LENGTH: 2171
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-291

Query Match 2.9%; Score 83; DB 5; Length 2171;
Best Local Similarity 20.0%; Pred. No. 58;
Matches 85; Conservative 55; Mismatches 152; Indels 132; Gaps 22;

QY 200 ELAKKADQVTFM--RPSYCLPMRQRTMDNE-----QTAKKAYI-----PTLFEAS 245
| | | | : | | | : | | | : | | | : |
Db 1130 QLAEKALNLSKMWGRKMSVQTPLRQFHGLNSDILMQLKKDLVWERYDLSAQLGELI 1189
| | | | : | | | : | | | : | | | : |

QY 246 RKSRIQFPV-----QAPSQVGI-FEVSPEQREAYFEELWEGAFNLAQCYREVVMVDKKAN 299
| | | | : | | | : | | | : | | | : |
Db 1190 RSPKNGRPLHKFIHQFPKVTLSAHVQPIRTVNLVNLVTPTDF-----LWDEKIH 1239
| | | | : | | | : | | | : | | | : |
```

```
QY 300 RLVTDFWAKTKRTRIV--NPAKR-----DLMAPELPY --- 330
| | | | : | | | : | | | : | | | : |
Db 1240 KYVEPFWI-----IVEDNDGEKILHHYFLLKKOYIDEDHTLHTFTVFEPLPOYFVR 1293
| | | | : | | | : | | | : | | | : |

QY 331 -----WFGTKR-SPLESDYIEMLDK--PSVEIVNLEQSPIVAVTKTG--VLLSDGS --- 376
| | | | : | | | : | | | : | | | : |
Db 1294 VVSDKWLGSSEIVLPVSFRHLILPEKYPPPTTELLDLOPLPVTALRNPNYEILYQDFKHPN 1353
| | | | : | | | : | | | : | | | : |

QY 377 -KRECDTIVLATGDFDSFTGSLTHMG-----LKNKH-GVD-----LKEVWKD 415
| | | | : | | | : | | | : | | | : |
Db 1354 VQTQVFTVLYNTNDNLVAAAPTSGKTKCAEFAILRNHHEGPDATMRVVYIAPLEAIAKE 1413
| | | | : | | | : | | | : | | | : |

QY 416 GISYTMGVFSGHPNAPFVATAQAP--TVLSNGPTIET-----OV 454
| | | | : | | | : | | | : | | | : |
Db 1414 QFRWEGKFGKGLRVLVVELTGETALDKLLEKQIITISPEKDALSRRWKORKVQOV 1473
| | | | : | | | : | | | : | | | : |

QY 455 DL-TADTTAKLEAEHATSVKSAQEAWSIMIAKMHNTLPLTDSW-----WTGG 505
| | | | : | | | : | | | : | | | : |
Db 1474 SLFIVDELHLITGGQHPVLEIVSRMYISSQV--INKIRIVALSTSLANAKDLGEMIGA 1531
| | | | : | | | : | | | : | | | : |

QY 506 NIPG 509
| | | | : | | | : | | | : |
Db 1532 SSHG 1535
| | | | : | | | : | | | : |

RESULT 11
US-09-573-655B-267
; Sequence 267, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573,655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 267
; LENGTH: 844
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-267

Query Match 2.8%; Score 82.5; DB 5; Length 844;
Best Local Similarity 20.2%; Pred. No. 17;
Matches 105; Conservative 65; Mismatches 200; Indels 151; Gaps 26;

QY 60 GARVDSET-----PFQNLNPEVWKDWTWSCRYPDQKELLSYVHCDKIRGLRKDVY 111
| | | | : | | | : | | | : | | | : |
Db 154 GKEVDSEALIAMSVGFVYSLTSEEIEANVSVIIGKQDQANYIVVRNH----- 201
| | | | : | | | : | | | : | | | : |

QY 112 FGAEVVDARYARDLGTWTVK-----TSAGH---VATAKYLIILATGLLRKHTPAL--PG 160
| | | | : | | | : | | | : | | | : |
Db 202 -----IIALRSNVSNLTFRDHALESIRAFKTLVDTAYNFLEHGYINFLGAPVIKEAK 256
| | | | : | | | : | | | : | | | : |

QY 161 LADFNGKVIHSSAWHEDFDAQGVAVIGAGATSIQIVQELAKKADQVTMFMRRPSYCLP 220
| | | | : | | | : | | | : | | | : |
Db 257 LRSFDG-----VEPPNVVVVAGLAGLVAARQLLSMGFRVLVLEGRDPRGGR 303
| | | | : | | | : | | | : | | | : |

QY 221 MRQRTMDRNEQTAWKA-----YYPTLFEASRKSRTGFPVQAPSVCIFE 263
| | | | : | | | : | | | : | | | : |
Db 304 VKTRMKGGDGVAMADVAGVSVLTGNGNPLGVLARQLGLPLHKVRIDICPLYLPNGELAD 363
| | | | : | | | : | | | : | | | : |

QY 264 VSPEQR-EAYFEELWEGAFNLAQCYREVVMVDKKAN-----RLVYDFWAKKT 310
| | | | : | | | : | | | : | | | : |
Db 364 ASVDSKIEASFKNLLDR-----VCKLRSMIEENKSVDPVPLGEALTFRLVYGV-AEDQ 416
| | | | : | | | : | | | : | | | : |

QY 311 RSRI-----VNPAKRDLMAPELPYVFGTKRSPLESYDEM-----LDKPSVEIVN 356
| | | | : | | | : | | | : | | | : |
Db 417 QERMLLDWHLANLEYANATLLGNLSMAYWD-----QDDPYEMGGDHCFIPGNGEIVH 469
| | | | : | | | : | | | : | | | : |
```

QY 357 --LEQSPI-----VAVTKTGVLVSDGSKR-BCD-----TI-----VLATGFDSTGSLTH 398
Db 470 ALAENLPIFYGSTVESRYSGNGLVVTGNKFEHCDMALCTVPLGLKKGSIERYPELPH 529
QY 399 -----MGLKNKHGVDLK-EVWKDGLSTYMGVSHGFPNA---FFVATAQAPTIVLS 444
Db 530 KKEAIOQLGFGLLKVKVAMLPFCNFWGEEIDTF-GRUTED-PSTRGGEFFLYSYSS--VS 585
QY 445 NGPTIITVDLIADTIATKLEAHATSVEATKSAQEAWSIM 485
Db 586 GGP-----LLVALVAGDAERFELTSPDTSVKRVQLIL 618

RESULT 12
US-10-102-806-557
; Sequence 557, Application US/10102806
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103PIC1
; CURRENT APPLICATION NUMBER: US/10/102,806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 557
; LENGTH: 565
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (57)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (71)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (75)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (82)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (118)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (120)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (552)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-102-806-557

Query Match 2.8%; Score 82; DB 6; Length 565;
Best Local Similarity 20.9%; Pred. No. 11;
Matches 95; Conservative 45; Mismatches 167; Indels 148; Gaps 23;

QY 25 SGISAVYKRLKRLNAKIFEGAPDFGGVWHNRYPCARVDSPTFPYQLNPIPEVKMDWTWS 84
Db 96 SGYRVHEELNLGLYKTRDGSXXG--WCW---PG-----S 127
QY 85 CRYPD-----KQELLSYVHCDKIRGL-----RKDYVFGAEVVDARYADLGTW 128
Db 128 AGYPDTNPTMRANWAMNPSYDNEGSAPNLFVWMDNPEPSFNGPEVTMLKDAQHYGGW 187
QY 129 ---TVKTSAG---HVATAXYLILATGLLRKHHTPALPGLADFNGLKVIHSSAWHEDFDAEG 182

Db 188 EHRDVHNIYGLVYHMTADGLRQRSGGMR-----PFVLARAFFAGSRQRFQAVWTGDMTAEW 244
QY 183 QRVAIVIGATSIQIVQELAKKADQVTFMRRPSCYCLPMQRQRTMDRNEQTAKKAYPTLF 242
Db 245 DHLKTSIPCLSLGLVGLSFCCAD-VCGFFKQKPEPELLVRWYMG-----AYQP-FF 294
QY 243 EASRKSRIQFPVQAPSVGIFVSPQREAYFEELWERGAFNPLACQYREYMDKKANRL- 301
Db 295 RAHAHLDTG-----RREPW-----LLPSQHNDIIRDALGORYS 327
QY 302 VYDFWAKKTRSRIVNPAKRDLMAPLEPPYVWFGTKRSPLESDDYYEMLDKP-SVEIVNLEQS 360
Db 328 LLPFW-----YLLYQAHREGI-PVMRPLW-----VQYPODVTTFNIDQ 366
QY 361 PIVAVTKTGVLVSDGSKRECDTIVLATGDFSFTGSLTHMGLKNKHGVDL-----KEWKD 415
Db 367 YLLG-----DALLVHPVSDS-----GARGVQVYLPQGGVEWYD 399
QY 416 GISTYMGVSHGFPNFAFVATAQAPTIVLSNGPTII 450
Db 400 -IQSYOK--HHGPOTLYLPVTLSSIPVFORGCTIV 431

RESULT 13
US-09-540-209B-6020
; Sequence 6020, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 6020
; LENGTH: 536
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-6020

Query Match 2.8%; Score 81; DB 5; Length 536;
Best Local Similarity 27.2%; Pred. No. 12;
Matches 40; Conservative 15; Mismatches 64; Indels 28; Gaps 8;

QY 414 KDGISTYMGVSHGFPNFAFV-----ATAQAPTIVLSNG-PTIETQ-VDLIADTIA 462
Db 176 KDDAGFVVGHASNGAAKALLAKVYVTMAGSAMSGVPIVYKGGNPNIFEPOPITHIAKTA 235
QY 463 KLEAEHATSVEATKSAQEAWSIMIAKNEHTLFP-LTDSWMTGGNIPGKATRALTPIGGI 521
Db 236 GYESFDPKAYVAL-ARDKAWEV---INEYTLFDNMDVWAIIGNRNKGHEHWAQAISGD 230
QY 522 ALY-EQICQEV-----ANWDG 537
Db 291 KDFGNTICQDYVGIFKEDGTMEGNWYG 317

RESULT 14
US-09-573-655B-117
; Sequence 117, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypepti
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573,655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 117
; LENGTH: 947
; TYPE: PRT

ORGANISM: Arabidopsis thaliana
US-09-573-655B-117

Query Match 2.8%; Score 81; DB 5; Length 947;
Best Local Similarity 23.2%; Pred. No. 27;
Matches 94; Conservative 47; Mismatches 157; Indels 108; Gaps 23;

QY 152 RKHTPALPLGLADP-----NGKVIHSS-----AWHEPFD-AEGQRVAVIGAGATSIQIVQE 200
DB 212 RKATPKGLKLRPIEAPDGTLVHDSYVGVGENAWDDDLTETTESGLKKTIGRNAR-----IQT 267
QY 201 LAKK-----ADQVTFMRPSPSYCLPMRQRTMDRNEQTAMKAYYPITLFEASRKS 248
DB 268 EAKKLSQDLGVSGETGDSVGNWRRLATWKEMLEKSEQLNSSAAKYVVEFDMKEVE 327
QY 249 RIGFPVQAPSVGEFVEPEQREAYFEELWERGAFNGLACQYREVVMVDKKANRLVYDFWAK 308
DB 328 K---SLREDVIGRTSTEGTRALWISKRWWR-----YRP-----KLPYTYFLQ 367
QY 309 KTRSRIV-----NPAKRDLM-----PLEPPYWFGTKRSPLESDDYEMLDKPSVEIVNL 357
DB 368 KLDSEVAAVVFEDLKLRYVTMKEGFPLE--YIVDI---PLDPYLPETICNAGVEVDLL 422
QY 358 EQSPIVAVTKTVLLSDGSK-----RECQTVLVTG-----FD-----SF---T 393
DB 423 QKQIHFMKVFIALPLGILILFMFIRESAMLLDITSKRFLYKKYQNLDFMAYAEFILPV 482
QY 394 GSLTHMGLNKHGVLDKEVWK--DGISTYMGVESHGFPPNAFF---VATAQAPTIVLSNGPT 448
DB 483 GDVSETKSMYKEVVLGVDWDLDELMIYM-----GNPMQYVEKDVAFVRG--VLLSGPP 535
QY 449 ILETQVDLIADTIAK---LEAEHATSVEATKSAQEAWSIMIAKNNE 491
DB 536 --CTGKTLFARTLAKESGLPFVFASGAETDSEKSG----AAKINE 575

Search completed: April 27, 2002, 07:58:35
Job time: 736 sec

ORGANISM: Arabidopsis thaliana
US-09-573-655B-117

Query Match 2.8%; Score 80.5; DB 5; Length 327;
Best Local Similarity 20.7%; Pred. No. 6.7;
Matches 67; Conservative 38; Mismatches 112; Indels 107; Gaps 17;

QY 255 QAFSVGIFEVSPQR--EAYFEELWERGAFNGLACQY-----REVMDKKANRL 301
DB 55 QSPMTGVLEAYPCQAGAIYGNVIEGKLTFFQMYLQNGEYGDKNREISLPVGTNNM 114
QY 302 VYDFWAKKTRSRIV--NPAKRDLMAP-----LEPPYWFGTKRSPLESDDYEMLD--- 348
DB 115 IY--WGTPKYEELIYSNPV---VVAPQITIGGDLSSQY--FGLRKVSADTYYPVFDLYVT 168
QY 349 -KPS---VEIVNLEQSPIVATK-----TGVLSDGSKRECDTIVLATGDFSFTGSLT 397
DB 169 VKPAHIGTIELSAAMQRVVAGLVIVKNNKNGILSS-----SIAGMEV 211
QY 398 HMGKLNKHGVLDKEVWKDGIYSTYMGVESHGFPPNAFFVATAQAPTIVLSNGPTIILETQVDLI 457

RESULT 15
US-09-540-209B-6436
; Sequence 6436, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FR
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 6436
; LENGTH: 327
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-6436

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 07:45:39 ; Search time 389.72 Seconds
(without alignments)
492.224 Million cell updates/sec

Title: US-09-882-694A-3
Perfect score: 2902
Sequence: 1 MSATSNRSGDCSVACDALLIV.....QICQEKVANNWGFVYLHAPC 545

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_Main:*

- 1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep.*
- 2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
- 3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
- 4: /cgn2_6/ptodata/2/paa/US080_COMB.pep.*
- 5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
- 6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
- 7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
- 8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
- 9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
- 10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
- 11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
- 12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
- 13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
- 14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
- 15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
- 16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
- 17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
- 18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
- 19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
- 20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
- 21: /cgn2_6/ptodata/2/paa/US097_COMB.pep.*
- 22: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
- 23: /cgn2_6/ptodata/2/paa/US099_COMB.pep.*
- 24: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
- 25: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
- 26: /cgn2_6/ptodata/2/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|------------------|
| 1 | 2902 | 100.0 | 545 | 17 | US-09-351-224-3 |
| 2 | 2902 | 100.0 | 545 | 17 | US-09-351-224E-3 |
| 3 | 2902 | 100.0 | 545 | 17 | US-09-351-823-3 |
| 4 | 2902 | 100.0 | 545 | 20 | US-09-677-488-3 |
| 5 | 2902 | 100.0 | 545 | 20 | US-09-677-488A-3 |
| 6 | 2902 | 100.0 | 545 | 20 | US-09-677-682-3 |
| 7 | 2902 | 100.0 | 545 | 20 | US-09-677-682A-3 |

| | | | | | |
|----|-------|-------|-----|----|----------------------|
| 8 | 2902 | 100.0 | 545 | 20 | US-09-677-682B-3 |
| 9 | 2902 | 100.0 | 545 | 22 | US-09-882-694-3 |
| 10 | 2902 | 100.0 | 545 | 22 | US-09-882-694A-3 |
| 11 | 833.5 | 28.7 | 529 | 19 | US-09-504-358-22 |
| 12 | 833.5 | 28.7 | 529 | 23 | US-09-954-314-22 |
| 13 | 833.5 | 28.7 | 529 | 26 | US-60-120-702-22 |
| 14 | 804 | 27.7 | 547 | 20 | US-09-648-004-10 |
| 15 | 796 | 27.4 | 553 | 19 | US-09-504-358-6 |
| 16 | 796 | 27.4 | 553 | 23 | US-09-954-314-6 |
| 17 | 796 | 27.4 | 553 | 26 | US-60-120-702-6 |
| 18 | 690 | 23.8 | 439 | 18 | US-09-417-507-40604 |
| 19 | 588.5 | 20.3 | 439 | 18 | US-09-417-507-26928 |
| 20 | 546.5 | 18.8 | 522 | 17 | US-09-328-352-6840 |
| 21 | 508 | 17.5 | 243 | 18 | US-09-417-507-40608 |
| 22 | 485.5 | 16.7 | 306 | 18 | US-09-417-507-27508 |
| 23 | 481.5 | 16.6 | 603 | 21 | US-09-732-615-2 |
| 24 | 481.5 | 16.6 | 603 | 26 | US-60-170-214-2 |
| 25 | 477.5 | 16.5 | 519 | 17 | US-09-328-352-7853 |
| 26 | 465.5 | 16.0 | 538 | 16 | US-09-252-991A-23068 |
| 27 | 453 | 15.6 | 525 | 17 | US-09-328-352-6683 |
| 28 | 444 | 15.3 | 516 | 17 | US-09-328-352-6474 |
| 29 | 436 | 15.0 | 528 | 23 | US-09-902-540-12229 |
| 30 | 398 | 13.7 | 219 | 20 | US-09-675-784A-8877 |
| 31 | 396 | 13.6 | 530 | 17 | US-09-328-352-7333 |
| 32 | 362.5 | 12.5 | 489 | 22 | US-09-888-320-2 |
| 33 | 354.5 | 12.2 | 482 | 22 | US-09-897-516-6981 |
| 34 | 354.5 | 12.2 | 482 | 26 | US-60-215-161-6981 |
| 35 | 346.5 | 11.9 | 543 | 16 | US-09-252-991A-19684 |
| 36 | 323 | 11.1 | 351 | 16 | US-09-252-991A-25097 |
| 37 | 300.5 | 10.4 | 507 | 17 | US-09-328-352-4291 |
| 38 | 290 | 10.0 | 253 | 18 | US-09-417-507-44096 |
| 39 | 280 | 9.6 | 312 | 18 | US-09-417-507-24881 |
| 40 | 262.5 | 9.0 | 558 | 17 | US-09-351-150A-7 |
| 41 | 249 | 8.6 | 343 | 16 | US-09-248-796-20883 |
| 42 | 249 | 8.6 | 343 | 26 | US-60-096-409-20883 |
| 43 | 247 | 8.5 | 335 | 18 | US-09-417-507-41815 |
| 44 | 244.5 | 8.4 | 364 | 16 | US-09-248-796-16427 |
| 45 | 244.5 | 8.4 | 364 | 26 | US-60-096-409-16427 |

ALIGNMENTS

RESULT 1
US-09-351-224-3
; Sequence 3, Application US/09351224
; GENERAL INFORMATION:
; APPLICANT: DuVick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fummonisin Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/351,224
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: flavin monooxygenase
US-09-351-224-3

| | | | | | | | |
|-----------------------|--------|--|-----------|------------|-----|--------|------|
| Query Match | 100.0% | Score | 2902; | DB | 17; | Length | 545; |
| Best Local Similarity | 100.0% | Pred. No. | 1.5e-300; | | | | |
| Matches | 545; | Conservative | 0; | Mismatches | 0; | Indels | 0; |
| Gaps | 0; | | | | | | |
| QY | 1 | MSATSNRSGDCSVACDALLIVGAGLSGISAVYKLRRLRLNAKIFEGAPDPGGVHNNRYPG | 60 | | | | |
| Db | 1 | MSATSNRSGDCSVACDALLIVGAGLSGISAVYKLRRLRLNAKIFEGAPDPGGVHNNRYPG | 60 | | | | |
| QY | 61 | ARVDSETPFYQLNTPYVWKDWTWSCRYPDOKELLSYVHHCCKIRGLRKDVYFGAEVVDAR | 120 | | | | |

Db 61 ARVDSPTFYQLNIPVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDYVFGAEVVDAR 120
QY 121 YARDLGTWTVKTSAGHVATAKYILILATGLLHRKHTPALPGADFNKGVIHSSAWHEDFDA 180
Db 121 YARDLGTWTVKTSAGHVATAKYILILATGLLHRKHTPALPGADFNKGVIHSSAWHEDFDA 180
QY 181 EGORVAVICAGATSIOIVQELAKKADQVTFMRRPSYCLPMRQRTMDRNEQTAWKAYPT 240
Db 181 EGORVAVICAGATSIOIVQELAKKADQVTFMRRPSYCLPMRQRTMDRNEQTAWKAYPT 240
QY 241 LFEASRKSRIIGFPVQAPSVGIFVSPQREAYFEELWERGAFNPLACQYREVVMVDKKANR 300
Db 241 LFEASRKSRIIGFPVQAPSVGIFVSPQREAYFEELWERGAFNPLACQYREVVMVDKKANR 300
QY 301 LVYDFWAKKTRSRIVNPAKRDLMAPLEPPYFVGTGKRSPLESDDYEMLDKPSVEIVNLEQS 360
Db 301 LVYDFWAKKTRSRIVNPAKRDLMAPLEPPYFVGTGKRSPLESDDYEMLDKPSVEIVNLEQS 360
QY 361 PIVAVTKTGVLSDGSKRECDTIVLATGFDSTGSLTHMGLKNKHGVDLKEVWKDGISTY 420
Db 361 PIVAVTKTGVLSDGSKRECDTIVLATGFDSTGSLTHMGLKNKHGVDLKEVWKDGISTY 420
QY 421 MGVSFSGHPNAPFVATAQAFTVLSNGPTIITETQVDLIADTIKLEAEHATSVEATKSAQE 480
Db 421 MGVSFSGHPNAPFVATAQAFTVLSNGPTIITETQVDLIADTIKLEAEHATSVEATKSAQE 480
QY 481 AWSIMIAKMNHTLPLTDSWMTGGNIPGKATRALTFIGGIALYEQICOEKVANWDGFDV 540
Db 481 AWSIMIAKMNHTLPLTDSWMTGGNIPGKATRALTFIGGIALYEQICOEKVANWDGFDV 540
QY 541 LHAPC 545
Db 541 LHAPC 545

RESULT 2

US-09-351-224E-3

; Sequence 3, Application US/09351224E

; GENERAL INFORMATION:

; APPLICANT: Duvick, Jon

; APPLICANT: Maddox, Joyce

; APPLICANT: Gilliam, Jacob

; APPLICANT: Folkerts, Otto

; APPLICANT: Crasta, Oswald R.

; TITLE OF INVENTION: Compositions and Methods for Fumonisin

; FILE REFERENCE: 5718-111

; CURRENT APPLICATION NUMBER: US/09/351,224E

; CURRENT FILING DATE: 1999-07-12

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 545

; TYPE: PRT

; ORGANISM: Exophiala spinifera

US-09-351-224E-3

Query Match 100.0%; Score 2902; DB 17; Length 545;
Best Local Similarity 100.0%; Pred. No. 1.5e-300;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSATSNRSGDCSVACDAIIIVGAGLSGISAVYKLRKRLNNAKIFEGAPDGGVWHNNRYPG 60
Db 1 MSATSNRSGDCSVACDAIIIVGAGLSGISAVYKLRKRLNNAKIFEGAPDGGVWHNNRYPG 60
QY 61 ARVDSPTFYQLNIPVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDYVFGAEVVDAR 120
Db 61 ARVDSPTFYQLNIPVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDYVFGAEVVDAR 120
QY 121 YARDLGTWTVKTSAGHVATAKYILILATGLLHRKHTPALPGADFNKGVIHSSAWHEDFDA 180
Db 121 YARDLGTWTVKTSAGHVATAKYILILATGLLHRKHTPALPGADFNKGVIHSSAWHEDFDA 180
QY 181 EGORVAVICAGATSIOIVQELAKKADQVTFMRRPSYCLPMRQRTMDRNEQTAWKAYPT 240
Db 181 EGORVAVICAGATSIOIVQELAKKADQVTFMRRPSYCLPMRQRTMDRNEQTAWKAYPT 240
QY 241 LFEASRKSRIIGFPVQAPSVGIFVSPQREAYFEELWERGAFNPLACQYREVVMVDKKANR 300
Db 241 LFEASRKSRIIGFPVQAPSVGIFVSPQREAYFEELWERGAFNPLACQYREVVMVDKKANR 300

QY 301 LVYDFWAKTRSRIVNPAKRDLMAPLEPPYWFQTKRSPLESDYYEMLDKPSVEIVNLEQS 360
Db 301 LVYDFWAKTRSRIVNPAKRDLMAPLEPPYWFQTKRSPLESDYYEMLDKPSVEIVNLEQS 360
QY 361 PIVAVTKGVLLSDGSKRECDTIVLATGDFSTGSLTHMGLKNKHGVLDKEVWKDGISTY 420
Db 361 PIVAVTKGVLLSDGSKRECDTIVLATGDFSTGSLTHMGLKNKHGVLDKEVWKDGISTY 420
QY 421 MGVSFSGFPNAPFVATAQAQPTVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQE 480
Db 421 MGVSFSGFPNAPFVATAQAQPTVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQE 480
QY 481 AWSIMIAKNEHTLPLDTSWMTGGNIPGKATRALTFIGGIALYEQICQEKVANWMDGFDV 540
Db 481 AWSIMIAKNEHTLPLDTSWMTGGNIPGKATRALTFIGGIALYEQICQEKVANWMDGFDV 540
QY 541 LHAPC 545
Db 541 LHAPC 545
RESULT 4
US-09-677-488-3
; Sequence 3, Application US/09677488
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fumonisin Detoxification
; FILE REFERENCE: 35718/204100 (5718-111A)
; CURRENT APPLICATION NUMBER: US/09/677,488
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: flavin monooxygenase, translation of fully spliced
US-09-677-488-3
Query Match 100.0%; Score 2902; DB 20; Length 545;
Best Local Similarity 100.0%; Pred. No. 1.5e-300;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSATSNRGCDCSVACDAIIVGAGLSGISAVYKLRKLRLNAKIFEGAPDFGCVWHNNRYPG 60
Db 1 MSATSNRGCDCSVACDAIIVGAGLSGISAVYKLRKLRLNAKIFEGAPDFGCVWHNNRYPG 60
QY 61 ARVDSETPFYQLNIPVWKDWTWSCRYPDQKELLSYVHHCDKIRGLKDVYFGAEVVDAR 120
Db 61 ARVDSETPFYQLNIPVWKDWTWSCRYPDQKELLSYVHHCDKIRGLKDVYFGAEVVDAR 120
QY 121 YARDLGTWTKTSAGHVATAKYLIILATGLLHRKHHTPALPLGLADPNKGVIHSSAWHEDFDA 180
Db 121 YARDLGTWTKTSAGHVATAKYLIILATGLLHRKHHTPALPLGLADPNKGVIHSSAWHEDFDA 180
QY 181 EGORVAVIGAGATSIQIVQELAKADQVTFMRRPSPCLPMRQRTMDRNEQTAWKAYPT 240
Db 181 EGORVAVIGAGATSIQIVQELAKADQVTFMRRPSPCLPMRQRTMDRNEQTAWKAYPT 240
QY 241 LFEASRSKRIGFPVQAPSVGIFVSPQREAYFEELWERGAFNFLACQYREVWVDKKNR 300
Db 241 LFEASRSKRIGFPVQAPSVGIFVSPQREAYFEELWERGAFNFLACQYREVWVDKKNR 300
QY 301 LVYDFWAKTRSRIVNPAKRDLMAPLEPPYWFQTKRSPLESDYYEMLDKPSVEIVNLEQS 360
Db 301 LVYDFWAKTRSRIVNPAKRDLMAPLEPPYWFQTKRSPLESDYYEMLDKPSVEIVNLEQS 360

QY 361 PIVAVTKGVLLSDGSKRECDTIVLATGDFSTGSLTHMGLKNKHGVLDKEVWKDGISTY 420
Db 361 PIVAVTKGVLLSDGSKRECDTIVLATGDFSTGSLTHMGLKNKHGVLDKEVWKDGISTY 420
QY 421 MGVSFSGFPNAPFVATAQAQPTVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQE 480
Db 421 MGVSFSGFPNAPFVATAQAQPTVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQE 480
QY 481 AWSIMIAKNEHTLPLDTSWMTGGNIPGKATRALTFIGGIALYEQICQEKVANWMDGFDV 540
Db 481 AWSIMIAKNEHTLPLDTSWMTGGNIPGKATRALTFIGGIALYEQICQEKVANWMDGFDV 540
QY 541 LHAPC 545
Db 541 LHAPC 545
RESULT 5
US-09-677-488A-3
; Sequence 3, Application US/09677488A
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 35718/204100
; CURRENT APPLICATION NUMBER: US/09/677,488A
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-677-488A-3

Query Match 100.0%; Score 2902; DB 20; Length 545;
Best Local Similarity 100.0%; Pred. No. 1.5e-300;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSATSNRGCDCSVACDAIIVGAGLSGISAVYKLRKLRLNAKIFEGAPDFGCVWHNNRYPG 60
Db 1 MSATSNRGCDCSVACDAIIVGAGLSGISAVYKLRKLRLNAKIFEGAPDFGCVWHNNRYPG 60
QY 61 ARVDSETPFYQLNIPVWKDWTWSCRYPDQKELLSYVHHCDKIRGLKDVYFGAEVVDAR 120
Db 61 ARVDSETPFYQLNIPVWKDWTWSCRYPDQKELLSYVHHCDKIRGLKDVYFGAEVVDAR 120
QY 121 YARDLGTWTKTSAGHVATAKYLIILATGLLHRKHHTPALPLGLADPNKGVIHSSAWHEDFDA 180
Db 121 YARDLGTWTKTSAGHVATAKYLIILATGLLHRKHHTPALPLGLADPNKGVIHSSAWHEDFDA 180
QY 181 EGORVAVIGAGATSIQIVQELAKADQVTFMRRPSPCLPMRQRTMDRNEQTAWKAYPT 240
Db 181 EGORVAVIGAGATSIQIVQELAKADQVTFMRRPSPCLPMRQRTMDRNEQTAWKAYPT 240
QY 241 LFEASRSKRIGFPVQAPSVGIFVSPQREAYFEELWERGAFNFLACQYREVWVDKKNR 300
Db 241 LFEASRSKRIGFPVQAPSVGIFVSPQREAYFEELWERGAFNFLACQYREVWVDKKNR 300
QY 301 LVYDFWAKTRSRIVNPAKRDLMAPLEPPYWFQTKRSPLESDYYEMLDKPSVEIVNLEQS 360
Db 301 LVYDFWAKTRSRIVNPAKRDLMAPLEPPYWFQTKRSPLESDYYEMLDKPSVEIVNLEQS 360
QY 361 PIVAVTKGVLLSDGSKRECDTIVLATGDFSTGSLTHMGLKNKHGVLDKEVWKDGISTY 420
Db 361 PIVAVTKGVLLSDGSKRECDTIVLATGDFSTGSLTHMGLKNKHGVLDKEVWKDGISTY 420

QY 421 MGVSFGFPNAPFVATAQAAPTIVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQ 480
|||||
Db 421 MGVSFGFPNAPFVATAQAAPTIVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQ 480
|||||
QY 481 AWSIMIAKMNHEHTLPPLTDSWMTGGNIPGKATRALTFIGGIALYEQICQEVANWMDGFDV 540
|||||
Db 481 AWSIMIAKMNHEHTLPPLTDSWMTGGNIPGKATRALTFIGGIALYEQICQEVANWMDGFDV 540
|||||
QY 541 LHAPC 545
|||||
Db 541 LHAPC 545
|||||
RESULT 6
US-09-677-682-3
; Sequence 3, Application US/09677682
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; TITLE OF INVENTION: Compositions and Methods for Fumonisin Detoxification
; FILE REFERENCE: 35718/204101 (5718-111B)
; CURRENT APPLICATION NUMBER: US/09/677,682
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: flavin monooxygenase, translation of fully spliced
; OTHER INFORMATION: CDNA
US-09-677-682-3

Query Match 100.0%; Score 2902; DB 20; Length 545;
Best Local Similarity 100.0%; Pred. No. 1.5e-300;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSATSNRSGDCSVACDAIIVGAGLSGISAVYKLRKRLNAKIFEGAPDFGGVWHNRYPG 60
|||||
Db 1 MSATSNRSGDCSVACDAIIVGAGLSGISAVYKLRKRLNAKIFEGAPDFGGVWHNRYPG 60
|||||
QY 61 ARVDSPTFYQLNIPVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDYVFGAEVVDAR 120
|||||
Db 61 ARVDSPTFYQLNIPVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDYVFGAEVVDAR 120
|||||
QY 121 YARDLGTWTVKTSAGHVATAKYLIILATGLLHRKHHTPALPLADFNKGVIHSSAWHEDFDA 180
|||||
Db 121 YARDLGTWTVKTSAGHVATAKYLIILATGLLHRKHHTPALPLADFNKGVIHSSAWHEDFDA 180
|||||
QY 181 EGORVAVIGAGATSIOIQVELAKKADQVTFMRRPSYCLPMRQRTMDRNEQTAWKAYPT 240
|||||
Db 181 EGORVAVIGAGATSIOIQVELAKKADQVTFMRRPSYCLPMRQRTMDRNEQTAWKAYPT 240
|||||
QY 241 LFEASRSKRIGFPVQAPSVGIFVSPQREAYFEELWEGAFNFLACQYREVWVDKKANR 300
|||||
Db 241 LFEASRSKRIGFPVQAPSVGIFVSPQREAYFEELWEGAFNFLACQYREVWVDKKANR 300
|||||
QY 301 LVYDFWAKKTRSRIVNPAKRDLMAPLEPPYWFCTKRSPLESDDYEMLDKPSVEIVNLEQS 360
|||||
Db 301 LVYDFWAKKTRSRIVNPAKRDLMAPLEPPYWFCTKRSPLESDDYEMLDKPSVEIVNLEQS 360
|||||
QY 361 PIVAVTKTGVLSDGSKRECDTIVLATGDSFTGSLTHMGLKNKHGVLDKEVWKDGISTY 420
|||||
Db 361 PIVAVTKTGVLSDGSKRECDTIVLATGDSFTGSLTHMGLKNKHGVLDKEVWKDGISTY 420
|||||
QY 421 MGVSFGFPNAPFVATAQAAPTIVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQ 480
|||||
Db 421 MGVSFGFPNAPFVATAQAAPTIVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQ 480
|||||

QY 481 AWSIMIAKMNHEHTLPPLTDSWMTGGNIPGKATRALTFIGGIALYEQICQEVANWMDGFDV 540
|||||
Db 481 AWSIMIAKMNHEHTLPPLTDSWMTGGNIPGKATRALTFIGGIALYEQICQEVANWMDGFDV 540
|||||
QY 541 LHAPC 545
|||||
Db 541 LHAPC 545
|||||
RESULT 7
US-09-677-682A-3
; Sequence 3, Application US/09677682A
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 35718/204101
; CURRENT APPLICATION NUMBER: US/09/677,682A
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-677-682A-3

Query Match 100.0%; Score 2902; DB 20; Length 545;
Best Local Similarity 100.0%; Pred. No. 1.5e-300;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSATSNRSGDCSVACDAIIVGAGLSGISAVYKLRKRLNAKIFEGAPDFGGVWHNRYPG 60
|||||
Db 1 MSATSNRSGDCSVACDAIIVGAGLSGISAVYKLRKRLNAKIFEGAPDFGGVWHNRYPG 60
|||||
QY 61 ARVDSPTFYQLNIPVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDYVFGAEVVDAR 120
|||||
Db 61 ARVDSPTFYQLNIPVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDYVFGAEVVDAR 120
|||||
QY 121 YARDLGTWTVKTSAGHVATAKYLIILATGLLHRKHHTPALPLADFNKGVIHSSAWHEDFDA 180
|||||
Db 121 YARDLGTWTVKTSAGHVATAKYLIILATGLLHRKHHTPALPLADFNKGVIHSSAWHEDFDA 180
|||||
QY 181 EGORVAVIGAGATSIOIQVELAKKADQVTFMRRPSYCLPMRQRTMDRNEQTAWKAYPT 240
|||||
Db 181 EGORVAVIGAGATSIOIQVELAKKADQVTFMRRPSYCLPMRQRTMDRNEQTAWKAYPT 240
|||||
QY 241 LFEASRSKRIGFPVQAPSVGIFVSPQREAYFEELWEGAFNFLACQYREVWVDKKANR 300
|||||
Db 241 LFEASRSKRIGFPVQAPSVGIFVSPQREAYFEELWEGAFNFLACQYREVWVDKKANR 300
|||||
QY 301 LVYDFWAKKTRSRIVNPAKRDLMAPLEPPYWFCTKRSPLESDDYEMLDKPSVEIVNLEQS 360
|||||
Db 301 LVYDFWAKKTRSRIVNPAKRDLMAPLEPPYWFCTKRSPLESDDYEMLDKPSVEIVNLEQS 360
|||||
QY 361 PIVAVTKTGVLSDGSKRECDTIVLATGDSFTGSLTHMGLKNKHGVLDKEVWKDGISTY 420
|||||
Db 361 PIVAVTKTGVLSDGSKRECDTIVLATGDSFTGSLTHMGLKNKHGVLDKEVWKDGISTY 420
|||||
QY 421 MGVSFGFPNAPFVATAQAAPTIVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQ 480
|||||
Db 421 MGVSFGFPNAPFVATAQAAPTIVLSNGPTIETQVDLIADTIAKLEAEHATSVEATKSAQ 480
|||||
QY 481 AWSIMIAKMNHEHTLPPLTDSWMTGGNIPGKATRALTFIGGIALYEQICQEVANWMDGFDV 540
|||||
Db 481 AWSIMIAKMNHEHTLPPLTDSWMTGGNIPGKATRALTFIGGIALYEQICQEVANWMDGFDV 540
|||||

QY 541 LHAPC 545
|||||
Db 541 LHAPC 545

RESULT 8

US-09-677-682B-3
; Sequence 3, Application US/09677682B
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 35718/204101
; CURRENT APPLICATION NUMBER: US/09/677,682B
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-677-682B-3

Query Match 100.0%; Score 2902; DB 20; Length 545;
Best Local Similarity 100.0%; Pred. No. 1.5e-300; Indels 0; Gaps 0;
Matches 545; Conservative 0; Mismatches 0;

QY 1 MSATSNRSGDCSVACDAIIVGAGLSGISAIVYKLRKLNKAKIFEGAPDFGGVWHNNRYPG 60
|||||
Db 1 MSATSNRSGDCSVACDAIIVGAGLSGISAIVYKLRKLNKAKIFEGAPDFGGVWHNNRYPG 60
|||||

QY 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLKRDVYFGEVVDAR 120
|||||
Db 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLKRDVYFGEVVDAR 120
|||||

QY 121 YARDLGTWTVKTSAGHVATAKYLILATGLLHKKHTPALPLADPNKGKVIHSSAWHEDFDA 180
|||||
Db 121 YARDLGTWTVKTSAGHVATAKYLILATGLLHKKHTPALPLADPNKGKVIHSSAWHEDFDA 180
|||||

QY 181 EGORVAVIGAGATSIQIVQELAKKADQVTFMRRPSYCLPMRQRTMDRNEQTANKAYPT 240
|||||
Db 181 EGORVAVIGAGATSIQIVQELAKKADQVTFMRRPSYCLPMRQRTMDRNEQTANKAYPT 240
|||||

QY 241 LFEASRKRIGFPVQAPSVGIFEVSPQREAYFEELWERGAFNPLACQYREVVMVDKKNR 300
|||||
Db 241 LFEASRKRIGFPVQAPSVGIFEVSPQREAYFEELWERGAFNPLACQYREVVMVDKKNR 300
|||||

QY 301 LVYDFWAKKTRSRIVNPAKROLMALEPPYWFGRKSPLESDDYEMLDKPSVEIVNLEQS 360
|||||
Db 301 LVYDFWAKKTRSRIVNPAKROLMALEPPYWFGRKSPLESDDYEMLDKPSVEIVNLEQS 360
|||||

QY 361 PIVAVTKTGVLSDGSKRECDTIVLATGFDSTGSLTHMGLKNKHGVDLKEVWKDGI 420
|||||
Db 361 PIVAVTKTGVLSDGSKRECDTIVLATGFDSTGSLTHMGLKNKHGVDLKEVWKDGI 420
|||||

QY 421 MGVSFGFPNPAFFVATAQAPTVLSNGPTIITQVDLIADTIAKLEAHSVEATKSAQE 480
|||||
Db 421 MGVSFGFPNPAFFVATAQAPTVLSNGPTIITQVDLIADTIAKLEAHSVEATKSAQE 480
|||||

QY 481 AWSIMIAKMEHTLPLDTSWMTGGNIPGKATRALTFIGGIALYEQICQEKVANWDGFDV 540
|||||
Db 481 AWSIMIAKMEHTLPLDTSWMTGGNIPGKATRALTFIGGIALYEQICQEKVANWDGFDV 540
|||||

QY 541 LHAPC 545
|||||

Db 541 LHAPC 545

RESULT 9

US-09-882-694-3
; Sequence 3, Application US/09882694
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 35718/208255
; CURRENT APPLICATION NUMBER: US/09/882,694
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-882-694-3

Query Match 100.0%; Score 2902; DB 22; Length 545;
Best Local Similarity 100.0%; Pred. No. 1.5e-300; Indels 0; Gaps 0;
Matches 545; Conservative 0; Mismatches 0;

QY 1 MSATSNRSGDCSVACDAIIVGAGLSGISAIVYKLRKLNKAKIFEGAPDFGGVWHNNRYPG 60
|||||
Db 1 MSATSNRSGDCSVACDAIIVGAGLSGISAIVYKLRKLNKAKIFEGAPDFGGVWHNNRYPG 60
|||||

QY 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLKRDVYFGEVVDAR 120
|||||
Db 61 ARVDSETPFYQLNIPEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLKRDVYFGEVVDAR 120
|||||

QY 121 YARDLGTWTVKTSAGHVATAKYLILATGLLHKKHTPALPLADPNKGKVIHSSAWHEDFDA 180
|||||
Db 121 YARDLGTWTVKTSAGHVATAKYLILATGLLHKKHTPALPLADPNKGKVIHSSAWHEDFDA 180
|||||

QY 181 EGORVAVIGAGATSIQIVQELAKKADQVTFMRRPSYCLPMRQRTMDRNEQTANKAYPT 240
|||||
Db 181 EGORVAVIGAGATSIQIVQELAKKADQVTFMRRPSYCLPMRQRTMDRNEQTANKAYPT 240
|||||

QY 241 LFEASRKRIGFPVQAPSVGIFEVSPQREAYFEELWERGAFNPLACQYREVVMVDKKNR 300
|||||
Db 241 LFEASRKRIGFPVQAPSVGIFEVSPQREAYFEELWERGAFNPLACQYREVVMVDKKNR 300
|||||

QY 301 LVYDFWAKKTRSRIVNPAKROLMALEPPYWFGRKSPLESDDYEMLDKPSVEIVNLEQS 360
|||||
Db 301 LVYDFWAKKTRSRIVNPAKROLMALEPPYWFGRKSPLESDDYEMLDKPSVEIVNLEQS 360
|||||

QY 361 PIVAVTKTGVLSDGSKRECDTIVLATGFDSTGSLTHMGLKNKHGVDLKEVWKDGI 420
|||||
Db 361 PIVAVTKTGVLSDGSKRECDTIVLATGFDSTGSLTHMGLKNKHGVDLKEVWKDGI 420
|||||

QY 421 MGVSFGFPNPAFFVATAQAPTVLSNGPTIITQVDLIADTIAKLEAHSVEATKSAQE 480
|||||
Db 421 MGVSFGFPNPAFFVATAQAPTVLSNGPTIITQVDLIADTIAKLEAHSVEATKSAQE 480
|||||

QY 481 AWSIMIAKMEHTLPLDTSWMTGGNIPGKATRALTFIGGIALYEQICQEKVANWDGFDV 540
|||||
Db 481 AWSIMIAKMEHTLPLDTSWMTGGNIPGKATRALTFIGGIALYEQICQEKVANWDGFDV 540
|||||

QY 541 LHAPC 545
|||||

Db 541 LHAPC 545

RESULT 10
US-09-882-694A-3
; Sequence 3, Application US/09882694A
; GENERAL INFORMATION:
; APPLICANT: Duivick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/882, 694A
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 09/351,224
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-882-694A-3

Query Match 100.0%; Score 2902; DB 22; Length 545;
Best Local Similarity 100.0%; Pred. No. 1.5e-300;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSATSNRSGCSCVACDAIIVGAGLSGISAVYKLRKLNKAKIFEGADPGGVVHNNRYPG 60
Db 1 MSATSNRSGCSCVACDAIIVGAGLSGISAVYKLRKLNKAKIFEGADPGGVVHNNRYPG 60
Qy 61 ARYDSETPFYQLNPEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDYVFGAEVVDAR 120
Db 61 ARYDSETPFYQLNPEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDYVFGAEVVDAR 120
Qy 121 YARDLGTWTWTKTSAGHVATAKYLILATGLLHRRKHTPALPGLADFNKGVHSSAHHEFDA 180
Db 121 YARDLGTWTWTKTSAGHVATAKYLILATGLLHRRKHTPALPGLADFNKGVHSSAHHEFDA 180
Qy 181 EGORVAVIGAGATSIQIQVBLAKADQVTFMRRPSCYCLPMRQRTMDRNEQTAWKAYTPT 240
Db 181 EGORVAVIGAGATSIQIQVBLAKADQVTFMRRPSCYCLPMRQRTMDRNEQTAWKAYTPT 240
Qy 241 LFASRSKRISGFPVQAPSVGIFVSPQREAYFEELWEGAFNPLACQYREVWVYDKKANR 300
Db 241 LFASRSKRISGFPVQAPSVGIFVSPQREAYFEELWEGAFNPLACQYREVWVYDKKANR 300
Qy 301 LVYDFWAKTRSRIVNPAKRDLMAPLEPPYWFGTKRSPLESDDYEMLDKPSVEIVNLEQS 360
Db 301 LVYDFWAKTRSRIVNPAKRDLMAPLEPPYWFGTKRSPLESDDYEMLDKPSVEIVNLEQS 360
Qy 361 PIVAVTKTGVLSDGSKRECDTIVLATGDSFTGSLTHMGLKNKHGVDLKEVWKGDGISTY 420
Db 361 PIVAVTKTGVLSDGSKRECDTIVLATGDSFTGSLTHMGLKNKHGVDLKEVWKGDGISTY 420
Qy 421 MGVSFSGFPNAPFVATAQPTVLSNGPTIITQVDLIADTIKLEAHATSVKATSAQE 480
Db 421 MGVSFSGFPNAPFVATAQPTVLSNGPTIITQVDLIADTIKLEAHATSVKATSAQE 480
Qy 481 AWSIMIAKNEHTLFLPLTDSWMTGGNIPGKATRALTFIGGIALYEQICQEKVANWGDGFDV 540
Db 481 AWSIMIAKNEHTLFLPLTDSWMTGGNIPGKATRALTFIGGIALYEQICQEKVANWGDGFDV 540
Qy 541 LHAPC 545
Db 541 LHAPC 545

RESULT 11
US-09-504-358-22
; Sequence 22, Application US/09504358

; GENERAL INFORMATION:
; APPLICANT: Rouviere, Pierre E.
; APPLICANT: Brzostowicz, Patricia C.
; TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF ADIPIC ACID INTERMED.
; FILE REFERENCE: BC1001 US NA
; CURRENT APPLICATION NUMBER: US/09/504,358
; CURRENT FILING DATE: 2000-02-15
; EARLIER APPLICATION NUMBER: 60/120,702
; EARLIER FILING DATE: 1999-February-19
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 22
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Brevibacterium sp HCU
US-09-504-358-22

Query Match 28.7%; Score 833.5; DB 19; Length 529;
Best Local Similarity 32.5%; Pred. No. 4.2e-79;
Matches 167; Conservative 107; Mismatches 237; Indels 3; Gaps 3;

Qy 5 SNRSGCSCVACDAIIVGAGLSGISAVYKLRKLNKAKIFEGADPGGVVHNNRYPGARVD 64
Db 11 AQANADETEVLDAIIVGGFSGVPVSDRLREDGFKYKVDAAAGGFGGIMWNCYPCGARTD 70
Qy 65 SETPPYQLNPEVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDYVFGAEVVDAR 124
Db 71 STGQIYQY 130
Qy 125 LGTWTWTKTSAGHVATAKYLILATGLLHRRKHTPALPGLADFNKGVHSSAHHE-DFDAEQ 183
Db 131 AKWTVRSSEGREQARAVIVATGFGAKPLYPNIEGLSDSFECECHHTARWPOGGLDWTGK 190
Qy 184 RVAVIGAGATSIQIQVBLAKADQVTFMRRPSCYCLPMRQRTMDRNEQTAWKAYTPTLFE 243
Db 191 RVVMTGASGIIQVQEAHAEHLTVFQRTNLPALPMRQRLSADDNDRYRENIEDREF 250
Qy 244 ASRSKRISGFPVQAPSVGIFVSPQREAYFEELWEGAFNPLACQYREVWVYDKKANRLVY 303
Db 251 IRDINSFAGDFYFQONAADTPEDERTAIYEKMWDEGGFPLWLGNGFQGLLTDEAHHFTY 310
Qy 304 DFWAKTRSRIVNPAKRDLMAPLEPPYWFGTKRSPLESDDYEMLDKPSVEIVNLEQSPV 363
Db 311 NFWRSKVHDRVKDPTAEMLAPATPHPGVKRPSLEQNYFDVYQNDVDLIDSNATPIT 370
Qy 364 AVTKTGVLSDGSKRECDTIVLATGDSFTGSLTHMGLKNKHGVDLKEVWKGDGISTYMGV 423
Db 371 RVLPGVETPDGV-ECVDVLTATGFDNNSGGINADIK-AGGQLLRDKWATGVDTYMGL 428
Qy 424 FSHGFPNAPFVATAQPTVLSNGPTIITQVDLIADTIKLEAHATSVKATSAQEAWS 483
Db 429 STHGFPNLMFLYGPSPSGFCNGTDFGGAPGDWADFLWLKDNKNGISFESTEEVEREWR 488
Qy 484 IMIAKNEHTLFLPLTDSWMTGGNIPGKATRALTF 517
Db 489 AHVDDIFVNSLFPKAKSWYGANVPKPAQMLNY 522

RESULT 12
US-09-954-314-22
; Sequence 22, Application US/09954314
; GENERAL INFORMATION:
; APPLICANT: Brzostowicz, Patricia C.
; APPLICANT: Rouviere, Pierre E.
; TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF ADIPIC ACID INTERMED.
; FILE REFERENCE: BC1001 US NA
; CURRENT APPLICATION NUMBER: US/09/954,314
; CURRENT FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/120,702
; PRIOR FILING DATE: 1999-February-19
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Microsoft Office 97

```

; SEQ ID NO 22
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Brevibacterium sp HCU
US-09-954-314-22

Query Match      28.7%; Score 833.5; DB 23; Length 529;
Best Local Similarity 32.5%; Pred. No. 4.2e-79;
Matches 167; Conservative 107; Mismatches 237; Indels 3; Gaps 3;

QY 5 SNSRGDCSVACDAIIVGAGLSGISAVYKLR-LRLNAKIFEGAPDGGVYHNNRYPGARVD 64
DB 11 AQAANADETEVLDAIIVGGFSGPVSDRLREDGFKVKVMDAAGGFGGIWNNCYPGARTD 70
QY 65 SETPFYQLNIPVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVYFGEVVDARYARD 124
DB 71 STGQIYQFOYKDLWKDFDEKELYPDFNGVREYFEYVDSQLDLSRDVTFTFAESCTWDDA 130
QY 125 LGTWTVKTSGHVATAKYLIATGLLHRKHTPALPGLADFNKGKVIHSSAWHE-DFDAEQG 183
DB 131 AKWTVRSSEGREQRARAVIVATGFGAKPLYPNIEGLDSFEGECHHTARWPGQGLDMTGK 190
QY 184 RVAVIGAGATSTQIWOELAKKADOVTFMRRPSCYCLPMRQRTMDRNEQTAWKAYPTLFE 243
DB 191 RVVVMGTGASGIQVIOEAAVAEHLTVFQRTPNLALPMRQRLSADDNDRYRENIEDRFQ 250
QY 244 ASRSKRIGFPVQAPSVEFEVSPQREAYFEELWGERAFNPLACQYREYVMDKKNRLVY 303
DB 251 IRDNSFAGDFYFIPQNAADTPEDERTAIYERKMDEGGFPLWLGNGFOLLTDEAANTFY 310
QY 304 DFWAKKTSRIVNPAKRDIMAPLEPPYFQKRSPLSDYYEMLDKPSVEIVNLEQSPIV 363
DB 311 NFWRKSVHDRVKDPKTAEMLAAPATPPHFGVKRPSLEQNYFDVYNODNVLLDSNATPIT 370
QY 364 AVTKTGVLSDGSKRECDTIVLATGDSFTGSLTHMGLKNKHGVDLKEVWKDGISTYMGV 423
DB 371 RVLPMGVETPDGVV-ECDVLVLATGDNSSGGINAIDIK-AGQQLLRDKWATGVDTYMGL 428
QY 424 FSHGFPNFAFVATAQAVTSLNSGPTIETQVDLIADTIAKLEAEHATSVEATKSAQEAWS 483
DB 429 STHGFPNLMFLYGPQSPGFCNGTDFGGAPGDMVADFLWLKNGISRFESTEVEVERWR 488
QY 484 IMIAKNEHTLPLDTSWMTGNNIPGKATRALTF 517
DB 489 AHVDDIFVNSLFPKAKSWYWGANGVPGKPAQMLNY 522

RESULT 13
US-60-120-702-22
; Sequence 22, Application US/60120702
; GENERAL INFORMATION:
; APPLICANT: ROUVIERE, PIERRE E.
; TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF
; TITLE OF INVENTION: ADIPIC ACID INTERMEDIATES
; FILE REFERENCE: BC-1001-PI
; CURRENT APPLICATION NUMBER: US/60/120,702
; CURRENT FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: MICROSOFT WORD VERSION 7.0A
; SEQ ID NO 22
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Brevibacterium sp hcu
US-60-120-702-22

Query Match      28.7%; Score 833.5; DB 26; Length 529;
Best Local Similarity 32.5%; Pred. No. 4.2e-79;
Matches 167; Conservative 107; Mismatches 237; Indels 3; Gaps 3;

QY 5 SNSRGDCSVACDAIIVGAGLSGISAVYKLR-LRLNAKIFEGAPDGGVYHNNRYPGARVD 64
DB 11 AQAANADETEVLDAIIVGGFSGPVSDRLREDGFKVKVMDAAGGFGGIWNNCYPGARTD 70
QY 65 SETPFYQLNIPVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVYFGEVVDARYARD 124
DB 71 STGQIYQFOYKDLWKDFDEKELYPDFNGVREYFEYVDSQLDLSRDVTFTFAESCTWDDA 130
QY 125 LGTWTVKTSGHVATAKYLIATGLLHRKHTPALPGLADFNKGKVIHSSAWHE-DFDAEQG 183
DB 131 AKWTVRSSEGREQRARAVIVATGFGAKPLYPNIEGLDSFEGECHHTARWPGQGLDMTGK 190
QY 184 RVAVIGAGATSTQIWOELAKKADOVTFMRRPSCYCLPMRQRTMDRNEQTAWKAYPTLFE 243
DB 191 RVVVMGTGASGIQVIOEAAVAEHLTVFQRTPNLALPMRQRLSADDNDRYRENIEDRFQ 250
QY 244 ASRSKRIGFPVQAPSVEFEVSPQREAYFEELWGERAFNPLACQYREYVMDKKNRLVY 303
DB 251 IRDNSFAGDFYFIPQNAADTPEDERTAIYERKMDEGGFPLWLGNGFOLLTDEAANTFY 310
QY 304 DFWAKKTSRIVNPAKRDIMAPLEPPYFQKRSPLSDYYEMLDKPSVEIVNLEQSPIV 363
DB 311 NFWRKSVHDRVKDPKTAEMLAAPATPPHFGVKRPSLEQNYFDVYNODNVLLDSNATPIT 370
QY 364 AVTKTGVLSDGSKRECDTIVLATGDSFTGSLTHMGLKNKHGVDLKEVWKDGISTYMGV 423
DB 371 RVLPMGVETPDGVV-ECDVLVLATGDNSSGGINAIDIK-AGQQLLRDKWATGVDTYMGL 428
QY 424 FSHGFPNFAFVATAQAVTSLNSGPTIETQVDLIADTIAKLEAEHATSVEATKSAQEAWS 483
DB 429 STHGFPNLMFLYGPQSPGFCNGTDFGGAPGDMVADFLWLKNGISRFESTEVEVERWR 488
QY 484 IMIAKNEHTLPLDTSWMTGNNIPGKATRALTF 517
DB 489 AHVDDIFVNSLFPKAKSWYWGANGVPGKPAQMLNY 522

RESULT 14
US-09-648-004-10
; Sequence 10, Application US/09648004
; GENERAL INFORMATION:
; APPLICANT: CHEN, QIONG
; APPLICANT: THOMAS, STUART
; TITLE OF INVENTION: BIOLOGICAL METHOD FOR THE PRODUCTION OF ADIPIC ACID AND
; TITLE OF INVENTION: INTERMEDIATES
; FILE REFERENCE: CL-1341-A
; CURRENT APPLICATION NUMBER: US/09/648,004
; CURRENT FILING DATE: 2000-04-25
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 10
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Acinetobacter sp.
US-09-648-004-10

Query Match      27.7%; Score 804; DB 20; Length 547;
Best Local Similarity 33.2%; Pred. No. 6.4e-76;
Matches 179; Conservative 100; Mismatches 230; Indels 30; Gaps 8;

QY 16 DAIVGAGLSGISAVYKLR-LRLNAKIFEGAPDGGVYHNNRYPGARVDSPTFY---- 70
DB 12 DAIVGGGFGGLYAVKKLRDELEUKVQAFDQATDVAGTGWYNNRYPCALSDTETHLYCYSW 71
QY 71 ---QNLNIPVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVYFGEVVDARYARDLGT 127
DB 72 DKELLQSLKIK-----KYVQGPDPVRKYLQOVAEKHDLKSKSYQNTAVQSAHYNEADAL 125
QY 128 WTVKTSAGHVATAKYLIATGLLHRKHTPALPGLADFNKGKVIHSSAWHEDFDAEQVAY 187
DB 126 WEVTEYGDKYTARFLITAGLLSAPNLPNIKGINQKGLHHTSRWPDVDSFEGKRGV 185

```

```

DB 11 AQAANADETEVLDAIIVGGFSGPVSDRLREDGFKVKVMDAAGGFGGIWNNCYPGARTD 70
QY 65 SETPFYQLNIPVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVYFGEVVDARYARD 124
DB 71 STGQIYQFOYKDLWKDFDEKELYPDFNGVREYFEYVDSQLDLSRDVTFTFAESCTWDDA 130
QY 125 LGTWTVKTSGHVATAKYLIATGLLHRKHTPALPGLADFNKGKVIHSSAWHE-DFDAEQG 183
DB 131 AKWTVRSSEGREQRARAVIVATGFGAKPLYPNIEGLDSFEGECHHTARWPGQGLDMTGK 190
QY 184 RVAVIGAGATSTQIWOELAKKADOVTFMRRPSCYCLPMRQRTMDRNEQTAWKAYPTLFE 243
DB 191 RVVVMGTGASGIQVIOEAAVAEHLTVFQRTPNLALPMRQRLSADDNDRYRENIEDRFQ 250
QY 244 ASRSKRIGFPVQAPSVEFEVSPQREAYFEELWGERAFNPLACQYREYVMDKKNRLVY 303
DB 251 IRDNSFAGDFYFIPQNAADTPEDERTAIYERKMDEGGFPLWLGNGFOLLTDEAANTFY 310
QY 304 DFWAKKTSRIVNPAKRDIMAPLEPPYFQKRSPLSDYYEMLDKPSVEIVNLEQSPIV 363
DB 311 NFWRKSVHDRVKDPKTAEMLAAPATPPHFGVKRPSLEQNYFDVYNODNVLLDSNATPIT 370
QY 364 AVTKTGVLSDGSKRECDTIVLATGDSFTGSLTHMGLKNKHGVDLKEVWKDGISTYMGV 423
DB 371 RVLPMGVETPDGVV-ECDVLVLATGDNSSGGINAIDIK-AGQQLLRDKWATGVDTYMGL 428
QY 424 FSHGFPNFAFVATAQAVTSLNSGPTIETQVDLIADTIAKLEAEHATSVEATKSAQEAWS 483
DB 429 STHGFPNLMFLYGPQSPGFCNGTDFGGAPGDMVADFLWLKNGISRFESTEVEVERWR 488
QY 484 IMIAKNEHTLPLDTSWMTGNNIPGKATRALTF 517
DB 489 AHVDDIFVNSLFPKAKSWYWGANGVPGKPAQMLNY 522

RESULT 14
US-09-648-004-10
; Sequence 10, Application US/09648004
; GENERAL INFORMATION:
; APPLICANT: CHEN, QIONG
; APPLICANT: THOMAS, STUART
; TITLE OF INVENTION: BIOLOGICAL METHOD FOR THE PRODUCTION OF ADIPIC ACID AND
; TITLE OF INVENTION: INTERMEDIATES
; FILE REFERENCE: CL-1341-A
; CURRENT APPLICATION NUMBER: US/09/648,004
; CURRENT FILING DATE: 2000-04-25
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 10
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Acinetobacter sp.
US-09-648-004-10

Query Match      27.7%; Score 804; DB 20; Length 547;
Best Local Similarity 33.2%; Pred. No. 6.4e-76;
Matches 179; Conservative 100; Mismatches 230; Indels 30; Gaps 8;

QY 16 DAIVGAGLSGISAVYKLR-LRLNAKIFEGAPDGGVYHNNRYPGARVDSPTFY---- 70
DB 12 DAIVGGGFGGLYAVKKLRDELEUKVQAFDQATDVAGTGWYNNRYPCALSDTETHLYCYSW 71
QY 71 ---QNLNIPVWKDWTWSCRYPDQKELLSYVHHCDKIRGLRKDVYFGEVVDARYARDLGT 127
DB 72 DKELLQSLKIK-----KYVQGPDPVRKYLQOVAEKHDLKSKSYQNTAVQSAHYNEADAL 125
QY 128 WTVKTSAGHVATAKYLIATGLLHRKHTPALPGLADFNKGKVIHSSAWHEDFDAEQVAY 187
DB 126 WEVTEYGDKYTARFLITAGLLSAPNLPNIKGINQKGLHHTSRWPDVDSFEGKRGV 185

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2002, 05:55:03 ; Search time 43.81 Seconds
(without alignments)
303.856 Million cell updates/sec

Title: US-09-882-694A-3
Perfect score: 2902
Sequence: 1 MSATSNRSGDCSVACDAIV.....QICQKVNWDGFDVLHAPC 545

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2.6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2.6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2.6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2.6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2.6/ptodata/2/1aa/PCTUS_COMB.pep.*
6: /cgn2.6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 202.5 | 7.0 | 520 | 4 | US-09-030-995-3 |
| 2 | 171.5 | 5.9 | 532 | 2 | US-08-676-841-8 |
| 3 | 170.5 | 5.9 | 532 | 2 | US-08-560-916-8 |
| 4 | 161.5 | 5.6 | 95 | 4 | US-09-655-270A-26 |
| 5 | 123.5 | 4.3 | 88 | 4 | US-09-655-270A-29 |
| 6 | 120.5 | 4.2 | 455 | 4 | US-09-030-995-2 |
| 7 | 118 | 4.1 | 92 | 4 | US-09-655-270A-28 |
| 8 | 116 | 4.0 | 73 | 4 | US-09-655-270A-32 |
| 9 | 100.5 | 3.5 | 491 | 4 | US-09-262-856A-5 |
| 10 | 99.5 | 3.4 | 544 | 1 | US-08-472-028A-8 |
| 11 | 99.5 | 3.4 | 544 | 2 | US-08-808-931-8 |
| 12 | 99.5 | 3.4 | 544 | 3 | US-08-808-323-8 |
| 13 | 99.5 | 3.4 | 544 | 3 | US-09-050-603A-8 |
| 14 | 99.5 | 3.4 | 544 | 3 | US-09-102-420B-8 |
| 15 | 99.5 | 3.4 | 544 | 4 | US-09-071-296-8 |
| 16 | 99.5 | 3.4 | 544 | 4 | US-09-196-268-8 |
| 17 | 99.5 | 3.4 | 544 | 4 | US-09-015-683-8 |
| 18 | 99.5 | 3.4 | 544 | 4 | US-09-191-998-8 |
| 19 | 99.5 | 3.4 | 544 | 4 | US-09-497-698-8 |
| 20 | 97.5 | 3.4 | 51 | 4 | US-09-655-270A-27 |
| 21 | 97 | 3.3 | 52 | 4 | US-09-655-270A-34 |
| 22 | 95.5 | 3.3 | 382 | 1 | US-07-783-705A-3 |
| 23 | 95 | 3.3 | 615 | 4 | US-09-297-937C-9 |
| 24 | 95 | 3.3 | 1276 | 4 | US-09-297-937C-13 |
| 25 | 93 | 3.2 | 711 | 1 | US-08-235-838-7 |
| 26 | 93 | 3.2 | 711 | 2 | US-08-465-473B-7 |
| 27 | 92.5 | 3.2 | 551 | 3 | US-08-886-886-15 |

| | | | | | | |
|----|------|-----|------|---|--------------------|-------------------|
| 28 | 92 | 3.2 | 549 | 3 | US-08-886-886-13 | Sequence 13, Appl |
| 29 | 91.5 | 3.2 | 921 | 4 | US-09-206-800-11 | Sequence 11, Appl |
| 30 | 91.5 | 3.2 | 1150 | 2 | US-08-589-756-3 | Sequence 3, Appl |
| 31 | 91.5 | 3.2 | 1150 | 4 | US-09-206-800-3 | Sequence 3, Appl |
| 32 | 91.5 | 3.2 | 1150 | 4 | US-09-206-898-3 | Sequence 3, Appl |
| 33 | 91 | 3.1 | 378 | 4 | US-09-199-637A-367 | Sequence 367, App |
| 34 | 89 | 3.1 | 1007 | 2 | US-08-551-459-4 | Sequence 4, Appl |
| 35 | 89 | 3.1 | 4150 | 4 | US-09-428-517-2 | Sequence 2, Appl |
| 36 | 86.5 | 3.0 | 1181 | 4 | US-09-206-898-23 | Sequence 23, Appl |
| 37 | 86 | 3.0 | 364 | 3 | US-08-817-926-51 | Sequence 51, Appl |
| 38 | 86 | 3.0 | 364 | 4 | US-09-203-893A-2 | Sequence 2, Appl |
| 39 | 86 | 3.0 | 364 | 4 | US-09-203-893A-4 | Sequence 4, Appl |
| 40 | 86 | 3.0 | 364 | 4 | US-09-203-893A-5 | Sequence 6, Appl |
| 41 | 86 | 3.0 | 364 | 4 | US-09-203-893A-8 | Sequence 8, Appl |
| 42 | 86 | 3.0 | 364 | 4 | US-09-203-893A-10 | Sequence 10, Appl |
| 43 | 86 | 3.0 | 364 | 4 | US-09-203-893A-12 | Sequence 12, Appl |
| 44 | 86 | 3.0 | 364 | 4 | US-09-203-893A-14 | Sequence 14, Appl |
| 45 | 86 | 3.0 | 364 | 4 | US-09-203-893A-16 | Sequence 16, Appl |

ALIGNMENTS

RESULT 1
US-09-030-995-3
; Sequence 3, Application US/09030995
; Patent No. 6235519
; GENERAL INFORMATION:
; APPLICANT: Wang, Yongzhao
; APPLICANT: Childs, John D.
; APPLICANT: Squires, Charles H.
; TITLE OF INVENTION: A Gene Involved in Thiophene
; TITLE OF INVENTION: Blotransformation from No. 6235519ardia Asteroides KGB1
; FILE REFERENCE: EBC97-05
; CURRENT APPLICATION NUMBER: US/09/030.995
; CURRENT FILING DATE: 1998-02-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-030-995-3

Query Match 7.0%; Score 202.5; DB 4; Length 520;
Best Local Similarity 21.9%; Pred. No. 1.4e-12;
Matches 108; Conservative 81; Mismatches 194; Indels 111; Gaps 22;
QY 19 IVGAGLSGISAVYKLRLLNAKIFEGAPDFGGVWHNRYPCARVDSETPFYQLNTPYVW 78
Db 7 VIGGVGSLSSIKCCVEGLEPVCFTDDIGGLRFOENP---EGRASIYKSVINTS 63
QY 79 KDMTWSCRYDPQKELLSVHHCDKIRGLR-----KDVYFGAENVVDARYARDL---G 126
Db 64 KEMMCFSDYPIPDHPVNPMMHNAQVLEYFRMYAKBFLLKYIRKTTVCVSKQPDFATSG 123
QY 127 TWTYKT-SAGIVATYAKY--LILATGLLRKHTP--ALPGLADFNKGVTHSSAMHEDFAE 181
Db 124 QMEVYTESGKEMNVFDGVNVCVCTGHTNAHLPLESPFCIEKFKQGYFHSRDYKKNPEGFT 183
QY 182 GORVAVTIGAGATSTQIVDELAKKADQVTMEMRRPSYCYLPMRQRTWDRNEQTAKWAYPTL 241
Db 184 GKRVIIGIGNSGDLAVEISQTAQVFLSTRGAWIL-----221
QY 242 FEASRKSRIGFPVQAPSVGIFEVSPQEAREYFEELWERCANFLACQYREVWVKKAN-R 300
Db 222 ---NRVDYGYPA-----DVLFSRLTTF--IW-KICQSLANKY----LEKKINQR 263
QY 301 LVYDFWAKKTRSRVNPNAKRLMAPLEPPYWFGRKSRPLESDYYEMLDKPSVEIVNL--E 358
Db 264 FDHEMFGLPKPKRAALS-----QHPTLND-----DLPNRIISGLVKV 299

QY 359 QSPIVAVTKGVLLSDGSKR-ECDTIVLATGDFDSFTGSLTHMGLKKNKHGVLDLKEVWKDGI 417
Db 300 KGNVKETETAAFEQSGREDDIDAVIFATGY-SFDEPPFLE-----DSYKVVANKI 349
QY 418 STYMGVSHGFPNAFFVATAQPTVLSNGPTIITQVDLIADTTAKLEAHATSVEATKS 477
Db 350 SLYKKVFP--PN-----LERPTLAIG-LIOLPLGAIN--PISELQGRWATQVFRGLK 396
QY 478 AQEANSIMIAKME 491
Db 397 TLPQSEMAEISK 410
RESULT 2
US-08-676-841-8
; Sequence 8, Application US/08676841
; Patent No. 5942426
; GENERAL INFORMATION:
; APPLICANT: Hayashi, Koji
; APPLICANT: Matsahi, Yasushi
; APPLICANT: Yabusaki, Yoshiyasu
; TITLE OF INVENTION: A HUMAN FLAVIN-CONTAINING MONOOXYGENASE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: CONCURRENTLY HERewith
; APPLICATION NUMBER: US/08/676,841
; FILING DATE: 18-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/560,916
; FILING DATE: 20-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 06-284902
; FILING DATE: 18-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakamura, Dean H.
; REGISTRATION NUMBER: 33,981
; REFERENCE/DOCKET NUMBER: Q39923
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 532 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-676-841-8

Query Match 5.9%; Score 171.5; DB 2; Length 532;
Best Local Similarity 20.4%; Pred. No. 2.9e-09;
Matches 92; Conservative 62; Mismatches 159; Indels 139; Gaps 19;

QY 19 IVGAGLSGISAVYKLRKLRLNAKIFEGAPDFGVYWHNR----- 57
Db 7 IIGAGVSGIASIRGLEGLEPTCFERSNDIGGLWKFSDHAEGRASIKSVFSNSSKEM 66
QY 58 --YPCARVDSTPTFYQLNIPVWKDWTWSCRYPQKELLSYVHHCKIRGLKRDYFGAE 115
Db 67 MCFDPDFPDFFPMFHN-----KIQEYIIAPAKEKLLKYIQPTFVSSVKNKHDPDFAT- 121
QY 116 VVDARYADLTGTWTVKISA-CHVATAKY--LILATGLLHRKHTP--ALPGLADFNKGVIH 170

Db 122 -----TGOWDVTTERDCKKESAVFDVAVMVCSGHHVYPLPKESFFGLNHFKGKCFH 172
QY 171 SSAMHEDFDEAGORVAVIGAGATSIQIQLAKKADQVMTMEMRRPSYCLPMRQRTMDRNE 230
Db 173 SRDKPEGVNGKRVLVVGVLGNSGCDIATELSRTAEQVMISSRSGSWM---SRVWDNG- 228
QY 231 QTAWKAYPTLFEASRKSRIQ----- 255
Db 229 -YPWDLMLVTRGCTFLKNNLPTAISDLVYVKOMNARFKHENYGLMPLNGVLKRPVFNDE 287
QY 256 APS-----VGIFEVSPEQRE-----AYFEE-----LWERG---AFNPL-----ACQY 289
Db 288 LPASILCGIVSVKPNVKEFTETSAIFEDGTIFEGIDCVIFATGYSFAYPFLDESIKSRN 347
QY 290 REVWYDK-----KANRLVYDF-----WAKKT-RSRIVNPAKRDLMA 324
Db 348 NEIILFKGVFPPLLEKSTIAVIGFVOSLGAAIPTVDLQSRWAAQVIKGTCTLPSPMEDMN 407
QY 325 PLEPPY-----WFGTKRSPLESDDYEMLDKPS 351
Db 408 DINEKMKRKNWFG-KSETIQDTIYVYMDLS 438
RESULT 3
US-08-560-916-8
; Sequence 8, Application US/08560916
; Patent No. 5928921
; GENERAL INFORMATION:
; APPLICANT: Hayashi, Koji
; APPLICANT: Matsahi, Yasushi
; APPLICANT: Yabusaki, Yoshiyasu
; TITLE OF INVENTION: A HUMAN FLAVIN-CONTAINING MONOOXYGENASE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/560,916
; FILING DATE: 20-NOV-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 06-284902
; FILING DATE: 18-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakamura, Dean H.
; REGISTRATION NUMBER: 33,981
; REFERENCE/DOCKET NUMBER: Q39923
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 532 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-560-916-8

Query Match 5.9%; Score 170.5; DB 2; Length 532;
Best Local Similarity 20.4%; Pred. No. 3.7e-09;
Matches 92; Conservative 62; Mismatches 159; Indels 139; Gaps 19;

QY 19 IVGAGLSGISAVYKRLRLNAKIFGAPDGGVWHNR-----57
Db 7 IIGAGVGLASIRSCLEGLEPTCFEKSNDIGGLWKFDAEGRASIIYKVSFNSKEM 66
QY 58 --YPGARVDETPFYQINPEVWKMWTSCRYDPQKELLSYVHHCDKIRGLRKYVFGAE 115
Db 67 MCFPDPFPDFFNFHNS-----KIOEYIIAFAKEKNLLKYIQKTFVSSVKNKHPDFAT- 121
QY 116 VVDARYARDLGTWVTKTSA--GHVATAKY--LILATGLLHRKHTP--ALPGADFNKGKVIH 170
Db 122 -----TGQWDVTERDGKESAVDAVMVCSGHVYVNPJPKSFPGNLHFKGKCFH 172
QY 171 SSAWHEDFDAEGORVAVICAGATSIQIOBELAKKADQVTFMFRPSPVCLPMQORTMDRNE 230
Db 173 SRDYKEPGVFNKRLVVLGNSGCDIATLSRTAQVMISRSQSWM---SRVWDNG- 228
QY 231 QTAWKAYYPTLEASKSRIG-----FPVQ 255
Db 229 -YPMOMLLVTRFTGLKNNLPTAISDLVYVKQMNARFKHENYGLMPLNGVLRKEPVFND 287
QY 256 APS---VGIFEVSPEORE-----AYFEE-----LWERG---AFNFL-----ACQY 289
Db 288 LPASILCGIVTKPNVKEFETSAIFEDGTIFEGIDCVIFATGYSFAYPFLDESIIKSRN 347
QY 290 REVMDK-----KANRLVYDF-----WAKKT-RSRIVNPAKRDLM 324
Db 348 NEILFKGVFPPLLEKSTIAVIGFVQSLGAAIPTVDLQSRWAAQVIKCTCTLPSMEDMMN 407
QY 325 PLEPPY-----WFGYKRSPLSDYYEMLDKPS 351
Db 408 DINERMEKRRKWF-KSETIQTQDIYVIMDELS 438
RESULT 4
US-09-655-270A-26
; Sequence 26, Application US/09655270A
; Patent No. 6329151
; GENERAL INFORMATION:
; APPLICANT: Rouviere, Pierre E.
; TITLE OF INVENTION: High Density Sampling of Differentially Expressed Prokaryotic m
; CURRENT APPLICATION NUMBER: US/09/655,270A
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/120,702
; PRIOR FILING DATE: 1999-February-19
; PRIOR APPLICATION NUMBER: 60/152,542
; PRIOR FILING DATE: 1999-September-03
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 26
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: microbial enrichment cultur
; OTHER INFORMATION: not one single organism
US-09-655-270A-26

Query Match 5.6%; Score 161.5; DB 4; Length 95;
Best Local Similarity 37.0%; Pred. No. 1.8e-09;
Matches 34; Conservative 17; Mismatches 40; Indels 1; Gaps 1;
QY 450 LETOVDLIADITAKLEAEHATSVEATKSAQEAWSIMIAKNEHTLPLTDSWNTGNTGPNIPG 509
Db 3 LETQVEWISDITVAYAEERNEIRIETPEAEBEWOTCTDIANATLFTRGDSWIFGANVPG 62
QY 510 KATRALTEFIGGIALYEQICQEKVA--NWDGFDV 540
Db 63 KKPSVLFILGLGNRYNVLGAWDSYRGFEL 94
RESULT 5

US-09-655-270A-29
; Sequence 29, Application US/09655270A
; Patent No. 6329151
; GENERAL INFORMATION:
; APPLICANT: Rouviere, Pierre E.
; TITLE OF INVENTION: High Density Sampling of Differentially Expressed Prokaryoti
; FILE REFERENCE: BC1011 US NA
; CURRENT APPLICATION NUMBER: US/09/655,270A
; CURRENT FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/120,702
; PRIOR FILING DATE: 1999-February-19
; PRIOR APPLICATION NUMBER: 60/152,542
; PRIOR FILING DATE: 1999-September-03
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 29
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: microbial enrichment cultur
; OTHER INFORMATION: not one single organism
US-09-655-270A-29
Query Match 4.3%; Score 123.5; DB 4; Length 88;
Best Local Similarity 34.1%; Pred. No. 1.7e-05;
Matches 29; Conservative 13; Mismatches 42; Indels 1; Gaps 1;
QY 457 IADTIAKLEAEHATSVEATKSAQEAWSIMIAKNEHTLPLTDSWNTGNTGNTGKATRALT 516
Db 3 ISDTIGYAEKNGVRAIETPEAEARMDBDLHRDRDNLTKGDSWIFGANIPGKTPSVLF 62
QY 517 FIGGIALYEQICQEKVAN--WDGFDV 540
Db 63 YLGGLRNYRAVLAEVATDGYRGFDV 87
RESULT 6
US-09-030-995-2
; Sequence 2, Application US/09030995
; Patent No. 6235519
; GENERAL INFORMATION:
; APPLICANT: Wang, Yongzhao
; APPLICANT: Childs, John D.
; TITLE OF INVENTION: A Gene Involved in Thiophene
; TITLE OF INVENTION: Biotransformation from No. 6235519ardia Asteroides KGB1
; FILE REFERENCE: EBC97-05
; CURRENT APPLICATION NUMBER: US/09/030,995
; CURRENT FILING DATE: 1998-02-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 455
; TYPE: PRT
; ORGANISM: No. 6235519ardia asteroides
US-09-030-995-2
Query Match 4.2%; Score 120.5; DB 4; Length 455;
Best Local Similarity 17.9%; Pred. No. 0.00057;
Matches 89; Conservative 70; Mismatches 150; Indels 189; Gaps 21;
QY 19 IVGAGLSGISAVYKRLRLNAK-----IPEGAPDFGCVWHNRYPGARVDSETPF 69
Db 7 VICAGSGLA--QLRAFOASAASKGADIPEIVCEKQANWGLMNYTWTGDLQYGE-PV 62
QY 70 YOLNIPVWK-----DWTWS-----CRYPDQKELLSYVHHCDKIRGLRKYDVF 112
Db 63 HGSWRYLWSNPKGEGLEFADYSFEHFGKQIASYPPRAVLFDYIEGRVKNAGYRDMWF 122
QY 113 GAQVVDARYARDLGTWVTKTSA-----GHVATAKY--LILATGLLHRKHHTPALPG 160

```

123 ENVV-----RLVTWDENTRKFTVTQDLPNDHCSYSEDVNVIVASGHFSTPNVPFPG 175
      |      |      |      |      |      |      |      |      |      |
161 LADPNGVKVIUSSAHWEDFDAGORVAVIGAGATSIQVQELAKKADOVTFMRRRSYCLIP 220
      |||::|: : : : | : ||
176 FDQNGRILHAHDFRDAREFIGHDVLLIGT-----SY----- 207
      |      |      |      |      |      |      |      |      |      |
221 MRQTDWRNEQTAWKAYYPFLFEASPKSRIGFPVQAPSVCIFEVSPQREAYFEELWERG 280
      |      |      |      |      |      |      |      |      |      |
208 -----SAEDIG-----SQCWKYG 220
      |      |      |      |      |      |      |      |      |      |
281 AFNELACOYREVVDVKKANLLVDFWAKKTRSRIVNPAKRDLMALEPPYWFCTKRSPL 340
      | : | : ||      |      |      |      |      |      |      |
221 ANSITNC-YR-----TKPMGY-----HW----- 237
      |      |      |      |      |      |      |      |      |      |
341 SDYVEMLDKPSVEIVNLEQSPIVAVTKGVLLSDGSKRCDTIVLATGF-DSPTGSLTHM 399
      | : | : || : || : : | : | : | : | : | : | : | : | : | : | :
238 PDNWE--EKLPLEKVHNTA-----TPDGSTKQIDAIICTGYKHHFFPLPDDL 285
      |      |      |      |      |      |      |      |      |      |
400 GLKNKHGVDLKVEKWKDGIS-----TYMGVFGSHGPNFAFFVATA-OAPTIVSNGFTIIE 451
      | : : : : | : | : | : | : | : | : | : | : | : | : | : | :
286 RLRTANRLATADLYK-GVAYVHNPALFYIGMQDQWFTENMFMDAQAWWARDVILGRIALPA 344
      |      |      |      |      |      |      |      |      |      |
452 TQVDLIADTTIAKLEAHA 469
      |      |      |      |      |      |      |      |      |      |
345 KGOEWADVAEARVAEDA 362
      |      |      |      |      |      |      |      |      |      |

```

```

RESULT 7
US-09-655-270A-28
; Sequence 28, Application US/09655270A
; Patent No. 6329151
; GENERAL INFORMATION:
; APPLICANT: Rouviere, Pierre E.
; TITLE OF INVENTION: High Density Sampling of Differentially Expressed Prokaryotic mR
; FILE REFERENCE: BC1011 US NA
; CURRENT APPLICATION NUMBER: US/09/655, 270A
; CURRENT FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/120,702
; PRIOR FILING DATE: 1999-February-19
; PRIOR APPLICATION NUMBER: 60/152,542
; PRIOR FILING DATE: 1999-September-03
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 28
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: microbial enrichment culture-
; OTHER INFORMATION: not one single organism
US-09-655-270A-28

```

```
Query Match      4.1%; Score 118; DB 4; Length 92;
Best Local Similarity 32.6%; Pred. NO. 7.1e-05;
Matches 30; Conservative 17; Mismatches 43; Indels 2; Gaps
QY 450 IETQVDLIADTIAKLEAHATSVSATKSQAQSASIMTAKNHHTLPPLTSWWTGNNPG 509
||||| :||| : : : ||| ||| ||| ||| |||
Db 1 IETQVEMISDTVPPTSATRSVRSPNPRS -RGGVADADLHRHREPTLFTRGDSVFGANVPG 59
QY 510 KATRALTIGGIATLYEQICQEKVA-NWDGFDV 540
60 KKPSVFLFYGLGLGNRYNNAGWADSRYRGGEL 91
Db
```

RESULT 8
US-09-655-270A-32
; Sequence 32, Application US/09655270A
; Patent No. 6329151
; GENERAL INFORMATION:
; APPLICANT: ROUVIERE, Pierre E.

Db 177 GI-EIDERTVSTGILSKVEPERLAI IGGGIIGLEMAVSAYARLGSKVTVI----- 227
QY 220 PMQRTMDRNEQAWKAYPTLFEASRKSIGFPVQAPSGVIGFEVSPQREAYFEELWER 279
Db 228 -----EFQNAIGAGMDAE----- 240
QY 280 GAFNFLACQREVMDKKNRLV---YDFWAKTTSRIYNPAKRDLMAPLEPPYWFGTK 335
Db 241 -----VAKSQKLLAKQGLDF---KLGTKVVKGERDGVVVKIEVDVKS GK 283
QY 336 RSPLESDDYEML---DRPSVEIVNLEQSPIVAVTKTGVLSDGSKRECDTIVLATGFSFT 393
Db 284 KSDLEADVLLVAIGRRPFTEGLNFEALGLEKDNKGRLLIIDDQFKTKHDHL-RVIGDVTFG 342
QY 394 GSITHMG-----LNKH-----GVDLKEVWKDGISTYMG 422
Db 343 PMLAHKAEIEGIAAEYIKKGGHVNYPNIPSYTHPEVAVYGLNEQLEQKGIKYKG 402
QY 423 VFSHGFPNFAFFVATAOAPTIVLSNGPTI-----IETQ-----VDLIADTIAKLEAEHATSVE 473
Db 403 KFP-----FIANSRAKTNMDTDFGVKFTIADAEQTVQLGVHIIIGPNAGEMIAEAGLAE 455
QY 474 ATKSAQE 480
Db 456 YGASTED 462

RESULT 10
US-08-472-028A-8
; Sequence 8, Application US/08472028A
; Patent No. 5767373
; GENERAL INFORMATION:
; APPLICANT: Ward, Eric R
; APPLICANT: Volrath, Sandra
; TITLE OF INVENTION: Manipulation of Protoporphyryrinen
; TITLE OF INVENTION: Oxidase Enzyme Activity in Eukaryotic Organisms
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/472,028A
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1748/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 544 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-472-028A-8

Query Match 3.4%; Score 99.5; DB 1; Length 544;
Best Local Similarity 20.1%; Pred. No. 0.13;
Matches 82; Conservative 47; Mismatches 115; Indels 163; Gaps 18;

QY 19 IVGAGLSGISAVYKLRKLRLNAKIFEGAPDFGVVHNNRYPCGARVDSETPPFYQLNIPYVW 78
Db 50 VVGAGVSLAAAYRURSGVNVTVFEAADRAGG-----KIRTNSEGGF-----VM 94
QY 79 KDWI-----WSC-----RYPQOKELLSYV----- 97
Db 95 DEGANTMEGEWEASRLDDGLQDKQYPSNQHKKRYIVKDGAPALIPSDPISLMKSSVL 154
QY 98 -----HHCDKIRGLKRDVYFGAEVVDARYDARDLGTWT 129
Db 155 STKSIALFFPPFLYKKANTRNSGKVSBEHLSVSGSFCERHFGREVVD--YFVD--PFV 210
QY 130 VKTSGHVATAKYLLIATGLLHRKHTPALPGLADENGKVIHSSAHMEDDFAEQORVAVIG 189
Db 211 AGTSAGD-----PESLSIRHAFPALNWLERYGYSVI-----VG 243
QY 190 AGATSIQIWOELAKKADQVTMFMRPSYCLPMRQR-----TMDRNEQOTAKAY--YPTLFE 243
Db 244 A-----ILSKIAKGD-----PVKTRHDSGKRRNRVVSFHHGMSQLIN 284
QY 244 ASRKS-----RIGPPVOA-----PSVGIFEVSPQREAYFEELWERGAFNFI----- 285
Db 285 ALHNEVGDDNVKLGTEVLSLACTFDGVFPALGRWSISVDSKDSCKDLASNOTFDAVIMTA 344
QY 286 -ACQYREVMDKKNRLVYDFWAKK---TRSRIVNPAKR-DLMAPIE 327
Db 345 PLSNVRRMKFTKGGAPVVDLPLRMDYLPISLMVTAFKDKDDVKKPLE 391

RESULT 11
US-08-808-931-8
; Sequence 8, Application US/08808931
; Patent No. 5939602
; GENERAL INFORMATION:
; APPLICANT: Volrath, Sandra
; APPLICANT: Johnson, Marie
; APPLICANT: Potter, Sharon
; APPLICANT: Ward, Eric
; APPLICANT: Heifetz, Peter
; TITLE OF INVENTION: DNA Molecules Encoding Plant
; TITLE OF INVENTION: Protoporphyryrinen Oxidase and Inhibitor-Resistant Mutants
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 5939602artis Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591-9005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/808,931
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/012,705
; FILING DATE: 28-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/013,612
; FILING DATE: 28-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/020,003
; FILING DATE: 21-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC 1847
; TELECOMMUNICATION INFORMATION:

```
; TELEPHONE: (919) 541-8587
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 544 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-808-931-8

Query Match      3.4%; Score 99.5; DB 2; Length 544;
Best Local Similarity 20.1%; Pred. No. 0.13;
Matches 82; Conservative 47; Mismatches 115; Indels 163; Gaps 18;

QY 19 IVGAGLSGISAVYKLRKLRLNAKIFGAPDFGGVWHNNRYPGARVDSETPFYQLNIPEVW 78
Db 50 VVGAGVSGLAAYRLRQSGVNVTVFEAADRAGG-----KIRTNSEGGF-----VW 94
QY 79 KDWT-----WSC-----RYPDOKELLSYV----- 97
Db 95 DEGANTWTEGEWEASRLIDDLGLQDKQOYPNSQHKRYIVKDGAPALIPSDPISLMKSSVL 154
QY 98 -----HHCDKIRGLRKDVYFGAEVVDARYARDLGTWT 129
Db 155 STKSKIALFPPEFLYKKANTRNSGKVSSEHLSSEVSGFCERHFGREVVD--YFVD--PFV 210
QY 130 VKTSAGHVATAKYLILATGLLHKRHTPALPGLADFNKGKVIHSSAWHEDFDAEGORVAVIG 189
Db 211 AGTSAGD-----PESLSIRHAFPALNWLRYKYSVI-----VG 243
QY 190 AGATSIQIVQELAKKADQVTMFMRRPSYCLPMRQR-----TMDRNEQTANKAY--YPTLFE 243
Db 244 A-----ILSKLAAGD-----PVKTRHDSGSKRRNRVVSFHHGGMQSLIN 284
QY 244 ASRKS-----RIGFPVQA-----PSVGIFEVSPEQREAYFEELWGERAFNFL---- 285
Db 285 ALHNEVGDDNVKLGTEVLSLACTFDGVPALGRWSISVDSKDSGDKDLASNOTFDAVIMTA 344
QY 286 -ACQYREVVDKANKRLVYDFWAKK--TRSRIVNPAKR-DLMAPLE 327
Db 345 PLSNVRRMKFTKGGAPVVDLFLPKMDYLPPLSLMVTAFKKDDVKKPLE 391

RESULT 12
US-08-808-323-8
; Sequence 8, Application US/08808323
; Patent No. 6018105
; GENERAL INFORMATION:
; APPLICANT: Johnson, Marie
; APPLICANT: Volrath, Sandra
; APPLICANT: Ward, Eric
; TITLE OF INVENTION: Promoters from Plant
; TITLE OF INVENTION: Protoporphyrogen Oxidase Genes
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSER: No. 6018105artis Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591-9005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/808,323
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/012,705
```

```
; FILING DATE: 28-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/013,612
; FILING DATE: 28-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/020,003
; FILING DATE: 21-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC 1846
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8587
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 544 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-808-323-8

Query Match      3.4%; Score 99.5; DB 3; Length 544;
Best Local Similarity 20.1%; Pred. No. 0.13;
Matches 82; Conservative 47; Mismatches 115; Indels 163; Gaps 18;

QY 19 IVGAGLSGISAVYKLRKLRLNAKIFGAPDFGGVWHNNRYPGARVDSETPFYQLNIPEVW 78
Db 50 VVGAGVSGLAAYRLRQSGVNVTVFEAADRAGG-----KIRTNSEGGF-----VW 94
QY 79 KDWT-----WSC-----RYPDOKELLSYV----- 97
Db 95 DEGANTWTEGEWEASRLIDDLGLQDKQOYPNSQHKRYIVKDGAPALIPSDPISLMKSSVL 154
QY 98 -----HHCDKIRGLRKDVYFGAEVVDARYARDLGTWT 129
Db 155 STKSKIALFPPEFLYKKANTRNSGKVSSEHLSSEVSGFCERHFGREVVD--YFVD--PFV 210
QY 130 VKTSAGHVATAKYLILATGLLHKRHTPALPGLADFNKGKVIHSSAWHEDFDAEGORVAVIG 189
Db 211 AGTSAGD-----PESLSIRHAFPALNWLRYKYSVI-----VG 243
QY 190 AGATSIQIVQELAKKADQVTMFMRRPSYCLPMRQR-----TMDRNEQTANKAY--YPTLFE 243
Db 244 A-----ILSKLAAGD-----PVKTRHDSGSKRRNRVVSFHHGGMQSLIN 284
QY 244 ASRKS-----RIGFPVQA-----PSVGIFEVSPEQREAYFEELWGERAFNFL---- 285
Db 285 ALHNEVGDDNVKLGTEVLSLACTFDGVPALGRWSISVDSKDSGDKDLASNOTFDAVIMTA 344
QY 286 -ACQYREVVDKANKRLVYDFWAKK--TRSRIVNPAKR-DLMAPLE 327
Db 345 PLSNVRRMKFTKGGAPVVDLFLPKMDYLPPLSLMVTAFKKDDVKKPLE 391

RESULT 13
US-09-050-603A-8
; Sequence 8, Application US/09050603A
; Patent No. 6023012
; GENERAL INFORMATION:
; APPLICANT: Volrath, Sandra
; APPLICANT: Johnson, Marie
; APPLICANT: Potter, Sharon
; APPLICANT: Ward, Eric
; APPLICANT: Helfetz, Peter
; TITLE OF INVENTION: DNA Molecules Encoding Plant
; TITLE OF INVENTION: Protoporphyrogen Oxidase
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6023012artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
```